LDK 2023

Cross-Lingual Transfer Learning for Misinformation Detection: Investigating Performance Across Multiple Languages

Oguzhan Ozcelik[1,2,†], Arda Sarp Yenicesu[2,†], Onur Yildirim[2,†],
Dilruba Sultan Haliloglu[2,†], Erdem Ege Eroglu[2,†], and Fazli Can[2]
[1]Aselsan Research Center, Ankara, Turkey
[2]Computer Engineering Department, Bilkent University, Ankara, Turkey
{oguzhan.ozcelik,sarp.yenicesu,o.yildirim,sultan.haliloglu}@bilkent.edu.tr,
ege.eroglu@ug.bilkent.edu.tr,
canf@cs.bilkent.edu.tr

Abstract

Detection of misinformation on social media requires human-annotated datasets to achieve truthful results. However, the annotation process is time-consuming due to the difficulty of labeling the veracity of the claims. Furthermore, most of the annotated misinformation detection datasets in the social media domain predominantly reside in English. To overcome this problem, we investigate the performance of cross-lingual transfer learning for misinformation detection across various languages, including, Arabic, Chinese, Turkish, and Polish. For this purpose, we analyze three different experimental setups on multilingual pre-trained language models in five natural languages (English, Arabic, Chinese, Turkish, and Polish). The results show that the multi-lingual mDeBERTa model can be applicable with fine-tuning in a widely-used language, i.e., English, and tested on a low-resource Turkish language with a successful recovery ratio, i.e., the metric shows the percentage of the recovered baseline score. For each model, we observe higher and more robust transfer ability between Polish and Arabic. Furthermore, it is possible to claim that contextual similarities outweigh language similarities, due to unsuccessful transfer learning ability between the English-Polish language pair.

1 Introduction

With the extensive use of social media, assessing the credibility of news has become a demanding task as the community is exposed to a substantial amount of information. Moreover, with the success of transformer-based auto-regressive models, it becomes challenging for a human reader to determine the reliability of the source of news (Hsu and Thompson, 2023). To overcome this issue, large language models (LLMs) become more popular to determine the veracity of a given news article (Kaliyar et al., 2021). However, it is challenging to develop a robust task-dependent LLM in low-resource languages due to the limitations of the training corpus. In this work, we will conduct detailed experiments to observe the cross-lingual transfer learning in the misinformation detection domain across various languages. Our study provides insight into which natural languages can be adapted to others, where the target domain limits the availability of an organized dataset.

Constructing a misinformation detection dataset is a challenging task as it requires human experts in the corresponding domain to annotate the disputed news (Shu et al., 2017a). Therefore, our experimental procedure employs multilingual pre-trained models to explore the transfer abilities of natural languages. The motivation of this study is to show how state-of-the-art approaches perform in low-resource languages when the source data is a widely-spoken language, i.e., English. Thus, we discuss the ways to choose a source language for a target language when the target language is limited in resources[1].

Misinformation detection can be performed on both noisy social media posts (Shu et al., 2017b) and well-written news articles (Wang, 2017). A common approach is training a classifier for a human-annotated dataset and predicting the veracity classes on a test collection. However, if a natural language has limited sources, the implementation and up-to-dateness of the proposed methods turn out to be an issue for that language.

1.1 Research Questions

To combat misinformation when there is a data limitation problem, we answer the following research questions:

[†]These authors contributed equally to this work

[1]During this study, we use the "low-resource language" term for the misinformation detection task. Although a language has limited resources in the misinformation detection task, it can be high-resourced for other natural language processing problems.

Table 1: **The available annotated misinformation detection datasets in English, Arabic, Chinese, Turkish, and Polish languages**. The referenced datasets are composed of social media (Twitter or Weibo) texts. (∗) Note that the table is not totally comprehensive. In other words, there may be some datasets that have been overlooked, especially in English.

Language	No.	Available Datasets*
English	17	(Kochkina et al., 2018), (Ma et al., 2016), (Derczynski et al., 2017), (Ma et al., 2017), (Shu et al., 2020), (Gorrell et al., 2019), (Nguyen et al., 2019), (Nguyen and Yu, 2021), (Dai et al., 2020), (Cui and Lee, 2020), (Dharawat et al., 2022) , (Li et al., 2020), (Patwa et al., 2021), (Alam et al., 2021), (Cheng et al., 2021), (Dadkhah et al., 2023), (Toraman et al., 2022a)
Arabic	3	(Haouari et al., 2020), (Alam et al., 2021), (Hadj Ameur and Aliane, 2021)
Chinese	1	(Yang et al., 2021)
Turkish	1	(Toraman et al., 2022a)
Polish	1	(Jarynowski, 2020)

RQ-1: Can we use widely-spoken high-resource language, such as English, as a source language in misinformation for low-resource target languages?

RQ-2: Which low-resource source language can be a better candidate for a high-resource target language in terms of transfer ability of misinformation detection task among the pairs of English, Chinese, Arabic, Turkish, and Polish?

1.2 Contributions

There are several studies conducted, including but not limited to cross-lingual data on fake news detection task (Arif et al., 2022; Du et al., 2021; Chu et al., 2021). However, there are a very few misinformation detection studies involving low-resource languages, such as Turkish (Toraman et al., 2022a) and Polish (Jarynowski, 2020). To the best of our knowledge, our study is the first to investigate the transfer ability across aforementioned languages in misinformation detection. Our contributions are the following:

- This is the first misinformation detection study that explores the transfer ability including Turkish and Polish languages.
- Our investigation aims to determine the most effective multilingual model for effectively transferring the task of misinformation detection across different languages.

The rest of the paper is organized as follows, in Section 2, we briefly introduce previous studies conducted in the area of misinformation detection and cross-lingual transfer learning. In Section 3, we formulate our problem in detail. Our approach to investigating the transfer ability of misinformation detection in various languages is given in Section 4. Section 5 describes the datasets we use in our experiments. In Section 6, we describe the experimental setup and then provide the results we obtain in Section 7. We discuss the experimental results in Section 8. Next, we provide limitations and ethical considerations in Section 9. Finally, Section 10 concludes the paper.

2 Related Work

We review previous works in terms of datasets, misinformation detection, and cross-lingual transfer learning studies.

2.1 Datasets

Table 1 summarises the incomplete list of datasets that can be used for misinformation detection in various domains, e.g., politics (Kochkina et al., 2018), public health (Cui and Lee, 2020), and so on. All of these datasets consist of social media posts, which resemble an informal way of presenting information. From Table 1, we observe that English covers the majority of the studies in the misinformation/disinformation area; hence, we decided to acknowledge English as a high-resource language as opposed to others (Arabic (Haouari et al., 2020), Chinese (Yang et al., 2021), Turkish (Toraman et al., 2022a), and Polish (Jarynowski, 2020)). Note that we also accept Arabic as a high-resource language for this study since there is more than one misinformation detection dataset in the Arabic language.

2.2 Misinformation Detection

Misinformation detection has become an important task, due to the ease of reaching and sharing content with the popularity of social media. There are different approaches to solving this detection problem. For instance, Helmstetter and Paulheim (2018) propose an ensemble method to predict fake news in a weakly supervised manner. Their ensemble model includes both traditional machine learning approaches like SVM (Vapnik, 1999), and Naive Bayes. De et al. (2021) utilize a transformer-based model, using BERT (Devlin et al., 2018) as the backbone, for multilingual fake news detection. Their dataset consists of news articles collected from various news websites with translated versions to low-resource languages such as Vietnamese. Monti et al. (2019) use a geometric deep-learning method to identify fake news in a dataset collected from Twitter, a widely-used social media platform. Graph neural networks are employed to distinguish fake news (Meyers et al., 2020). Social contexts are also used as a supportive feature for news content in a transformer-based architecture (Raza and Ding, 2022).

2.3 Cross-lingual Transfer Learning

Limited resources in some languages for a specific task, such as misinformation detection, require the emergence of cross-lingual studies. Probabilistic methods for cross-lingual information retrieval are investigated (Nie et al., 1999; Xu et al., 2001). A recurrent neural network-based approach is utilized to investigate multilingual analysis for limited data (Can et al., 2018). Moreover, Sun et al. (2021) employ a multilingual response generation layer and a cross-lingual knowledge retrieval layer to handle the language barrier in the context of the conversation. Besides, studies based on transfer learning in terms of few-shot learning are carried out to overcome the limited data problem (Hardalov et al., 2022).

Some studies utilize additional extracted features from external multi-lingual sources. Wen et al. (2018) utilizes an approach for rumor verification, employing multimedia content and external information in other news platforms. They achieve good performance with the use of extracted features. Dementieva and Panchenko (2021) propose a feature called "cross-lingual evidence" to be utilized in fake news identification. This feature is based on the idea "if a news is true, the facts mentioned in different languages should be identical". They report that the state-of-art models that use this feature perform better than their default versions. (Hammouchi and Ghogho, 2022) propose a framework for fake news detection employing external pieces of evidence searched by the web to verify the veracity of the news in multilingual datasets.

3 Problem Formulation

Suppose we have a misinformation dataset in target language $F_T = \{(N_i^T, L_i^T)\}_i^{|F_T|}$ with $|F_T|$ microblog-veracity pairs, where for all i, N_i^T refers to a tweet with veracity label L_i^T. The veracity, L_i^T, represents whether a microblog includes true information or false information as a binary variable (Eq. 1).

$$L_i^T = \begin{cases} 1 & \text{if } N_i^T \text{ includes true claim} \\ 0 & \text{if } N_i^T \text{ includes false claim} \end{cases} \tag{1}$$

We also have a collection of social media datasets, C_S, in other source languages:

$$C_S : [F_1 = \{(N_i^1, L_i^1)\}_i^{|F_1|}, \ldots, F_z = \{(N_i^z, L_i^z)\}_i^{|F_z|}]_\gamma^{|C_S|} \tag{2}$$

In Eq. 2, $|C_S|$ refers to the number of available misinformation detection datasets in other source languages we accessed, and each F refers to a dataset in other source languages. Each dataset, similar to the F_T consists of microblog-veracity pairs. γ is used for indexing the datasets in the C collection.

We will have a multilingual model set, $H = \{\{h(N)\}_m^{|C_S|+1}\}_k^K$ which has $K \times (|C_S| + 1)$ pre-trained models. Each $h(N)$ represents a multilingual language model focusing on one of the source languages or the target language while using a pre-trained multilingual model, e.g., mBERT (Devlin et al., 2018). For the target language and other languages, there are $|C_S| + 1$ models (There are $|C_S|$ source languages and 1 target language.), and for each of them there are K different multilingual model architecture, i.e. $K = 3$ for mBERT (Devlin et al., 2018), XLM-R (Conneau et al., 2019), and mDeBERTa (He et al., 2020). For the F_T and $C_{S\gamma}$ for all γ, $H = \{h(N)\}_m^{|C_S|+1}$ will be fine-tuned using aforementioned pre-trained multilingual models in source languages which is the language used in F_M during the fine-tuning of the h_m.

Given F_T, C_S, and H, we want to find cross-lingual transfer ability on misinformation detection in the target language. To find this transfer ability,

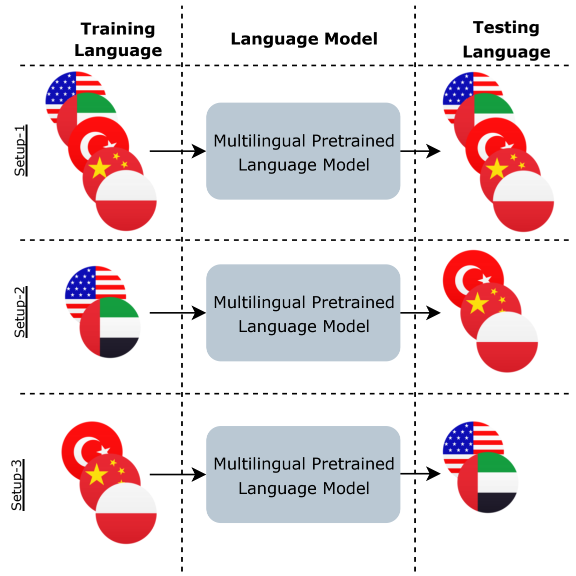

Figure 1: The illustration of our experimental methodology. (Setup-1) shows when a model is trained and tested in the same language for a specific task. (Setup-2) indicates when the language is crossed, i.e., training on a widely-used high-resource language (i.e., English or Arabic) and tested on low-resource languages. (Setup-3) simply represents when the model is trained and tested on low-resource, and high-resource languages, respectively.

first, we will evaluate h_t on F_T, the target dataset, e.g., CHECKED (Yang et al., 2021) if the target language is Chinese and achieve an F1 score, $F1_{\text{Target}}$. Then, we will repeat the same evaluation for all h_γ, where $h_\gamma \neq h_t$ on F_T and achieve a separate F1 score $F1_\gamma$, where $h_{t'}$ is fine-tuned using $C_{S\gamma}$. To evaluate the transfer ability of a language model, we employ relative zero-shot transfer ability (Turc et al., 2021) and call it "recovery ratio" following the study (Toraman et al., 2022b). We use the recovery ratio between the target language and the remaining languages from the C_S collection. The recovery ratio is formulated as in Eq. 3.

$$\text{Recovery Ratio}_\gamma = \frac{F1_\gamma}{F1_{\text{Target}}} \quad (3)$$

Finally, we will use these Recovery Ratio$_\gamma$ scores to compare and analyze the transfer learning ability of each source language in C_S to a target language.

4 Method

We investigate the transfer learning ability across five different languages, namely English, Chinese, Arabic, Turkish, and Polish. Particularly, we conduct analysis on a single NLP task, namely, misinformation detection. In order to find which language is a better choice when language transfer is required, we fine-tune pre-trained multilingual mBERT (Devlin et al., 2018), XLM-R (Conneau et al., 2019) and mDeBERTa (He et al., 2020) models in source languages, and predict the truthfulness of tweets (True or False) in a target language. The performance of the language transfer ability is evaluated on models via recovery ratio over baselines, where the baselines are the models fine-tuned and tested on the same language. In other words, we assume that the best performance occurs when the source and target language are the same. Thus, we use the baseline score as the denominator in Eq. 3.

We provide an illustration (see Figure 1), to explain our methodology for the experimental procedure. When a multilingual model is fine-tuned and tested in the same language, it yields promising results. However, for low-resource languages, such as Turkish, there are a few available data collections for specific problems, e.g., misinformation detection. This motivates cross-language studies to explore which widely spoken language can fit into a language if there is a lack of data collection in that language.

This methodology provides us an opportunity to empirically find the ability to transfer information from a high-resource source language to a low-resource target language while giving some valuable insights about hidden transfer mechanisms such as geopolitical influence on a language, shared vocabulary between languages, the impact of an alphabet on a language, and contextual similarities regardless of language differences.

5 Dataset

In this study, we use the English and Turkish microblogs from the splits of the MiDe-22 dataset (Toraman et al., 2022a), Chinese from the CHECKED dataset (Yang et al., 2021), Arabic from the AraCOVID19-MFH dataset (Hadj Ameur and Aliane, 2021) and Polish from Andrzej's dataset (Jarynowski, 2020). MiDe-22 is a tweet collection of misinformation domains, including various topics such as the Russo-Ukraine War, COVID-19, refugees, and so on, while CHECKED, AraCOVID19-MFH and Andrzej's only contain microblogs about COVID-19. For all datasets, we only use the true and false labeled social media posts in our experiments. The main statistics of the

Table 2: The main statistics of the datasets used in this study. The values are microblog counts for True labeled and False labeled microblogs.

Languages	English (🇺🇸)		Chinese (🇨🇳)		Arabic (🇦🇪)		Turkish (🇹🇷)		Polish (🇵🇱)	
Datasets	MiDe-22		CHECKED		AraCOVID19-MFH		MiDe-22		Andrzej's	
Splits	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
True	576	151	1,408	352	320	80	533	136	377	95
False	1,381	348	276	68	1,609	402	1,379	353	84	21
Total	1,957	499	1,684	420	1,929	482	1,912	489	461	116

datasets used in this study are given in Table 2.

6 Experimental Approach

In this study, we first define three experimental procedures. Then we utilize different multilingual pre-trained language models. We provide the details in the following sections.

6.1 Experimental Procedure

The experimental procedure consists of three types of setup (see Figure 1):

Setup-1 : When a model is trained and tested in the same language. (e.g., English → English)

Setup-2 : When a model is trained on a widely-used source language and tested on a low-resource language. (e.g., English → Turkish)

Setup-3 : When a model is trained on a low-resource source language and tested on high-resource languages. (e.g., Polish → Arabic)

In order to obtain a reference point for the recovery ratio metric, we construct "Setup-1". We assume that if a language model is trained and tested on the same language, its score is the maximum reference point to be achieved. Then, we implement "Setup-2" to answer **RQ-1**. Next, we use "Setup-3" for **RQ-2**. In order to investigate the better source language candidate, and transfer ability across languages, we evaluate recovery ratio metrics by employing the results of "Setup-1".

6.2 Language Models

We utilize three different multilingual pre-trained language models. The motivation behind choosing multilingual models is to have language knowledge of our studied languages in the pretraining corpus of these models. Thus, we can observe whether a specific task (in this study, the task is misinformation detection) can be learned via these models. The models are the following:

mBERT: BERT (Devlin et al., 2018) (Bidirectional Encoder Representations from Transformers) architecture serves as the foundation for the multilingual model known as mBERT. BERT was trained using Wikipedia and the Book Corpus dataset, which includes more than 10,000 books of various genres. To learn embedded representations of texts in many languages, this model is trained in a broad range of languages. mBERT can be used to process texts in several languages and for tasks like classification and translation because it supports multiple languages.

XLM-R: Cross-lingual Language Model - RoBERTa is what the acronym XLM-R (Conneau et al., 2019) stands for. A sizable pre-training dataset that included numerous huge, multilingual texts were used to train this model. Indeed, 100 languages from 2.5TB of filtered CommonCrawl data were used as its pre-training material. In order to learn embedded representations of multilingual texts, XLM-R employs an unsupervised learning technique. This makes it possible to identify semantic connections and commonalities across several languages.

mDeBERTa: Multilingual Decoding-enhanced BERT with Disentangled Attention is referred to as mDeBERTa (He et al., 2020). This model improves the BERT and RoBERTa (Zhuang et al., 2021) models using disentangled attention and enhanced mask decoder.

6.3 Experimental Setup

During the experiments, we use Hugging Face (Wolf et al., 2020) library to fine-tune Transformer-based language models. We choose learning rate 5e-5, batch size 16, the number of epochs 10, and maximum sequence length 128, following the study (Toraman et al., 2022a). During the training of the models, we employ an NVIDIA RTX A400. We use stratified five-fold cross-validation where the

Table 3: **Experimental results of Setup-1.** Column notations for metrics: precision (P), recall (R), and weighted F1 score (F1). Five-fold average precision, recall, and weighted F1 scores are reported.

Models	mBERT			XLM-R			mDeBERTa		
Datasets/Metrics	P	R	F1	P	R	F1	P	R	F1
MiDe-22-EN	0.879	0.880	0.879	0.724	0.806	0.758	0.884	0.881	**0.882**
AraCOVID19-MFH	0.998	0.998	**0.998**	0.997	0.997	0.997	0.997	0.997	0.997
CHECKED	0.991	0.991	0.991	0.996	0.996	**0.996**	0.996	0.996	**0.996**
MiDe-22-TR	0.894	0.895	0.894	0.885	0.886	0.885	0.902	0.901	**0.901**
Andrzej's	0.771	0.790	0.771	0.809	0.834	**0.794**	0.770	0.820	0.787

statistics of train and test splits are given in Table 2.

7 Experimental Results

We report the results obtained for Setup-1 in Table 3. Out of three multilingual language models, mDeBERTa produces higher F1 scores in English, Chinese, and Turkish datasets. On the other hand, mBERT performs better in Arabic, and XLM-R does it in the Polish language. The results are very high for Chinese and Arabic, with around 99% F1 scores. This is possible because these datasets are specifically on one topic, i.e., COVID-19. However, the Polish dataset is also in the COVID-19 domain but the models perform lower in Polish when compared to Chinese and Arabic. We may claim that Chinese and Arabic datasets are easier to detect misinformation possibly having biased patterns in texts.

From Table 4, we observe gray-highlighted cells, which are the average of weighted F1 scores on five-fold splits when source and target language are the same, i.e., Setup-1. For **RQ-2**, it can be seen that the mBERT model produces the highest score when it is trained in Arabic (a well-resourced language) and tested in Polish (a low-resource language) with a 95% recovery ratio. Similarly, the mDeBERTa achieves the highest score for the Arabic-Polish pair. For RQ-3, the XLM-R model produces the highest recovery ratio, 95%, with the Turkish-English pair. The rest of the experimental results are given in Section 4.

8 Discussion

In our studies, we use five languages from four different language families: Altaic (Turkish), European (English and Polish), Zhou (Chinese), and Sámi (Arabic). This separation gives us a fair ground for our experiments. In Table 4, we observe that the transfer ability from English to Turkish is higher than in any other source language. On average, we achieve an 84% recovery ratio for this transformation which suggest that English can be used as a source language for the Turkish language in a task-oriented setting, (**RQ1**). However, the transformation from English (as a high-resource language) to other low-resource languages except Turkish is not successful, and we arguably claim that this difference is due to contextual differences between the datasets used for the study where the datasets used for English and Turkish languages combined similar topics from the Russo-Ukraine War, COVID-19, refugees, etc., while others only focus on COVID-19, (**RQ1**). Moreover, even though Polish and English are in the same language family, the transfer performance between these two languages is low compared to some other pairs that contain Polish and English as either the target or the source. The reason behind these relatively lower scores between Polish and English can be due to the context of the data which supports our previous claim.

On the other hand, relatively lower results can be observed in the transfer ability of Arabic and Turkish, even though there are a lot of borrowed words. Another observation is the good transfer ability of Arabic to Chinese and vice versa. Since the Arabic and Chinese datasets both contain social media posts only about COVID-19, the performances of all models are better when these two languages are used as the source and the target languages. This also clearly shows that the domain of the data is essential and has an impact on the performance. This claim can be supported by the transfer ability performance from the Turkish language to the English language, where this transformation achieved an 88.3% recovery ratio on average of three models by utilizing similar misinformation domains. To conclude, if the domain of the data is similar, any low-resource language can be used as a source

Table 4: **The results of cross-lingual fake news experiments (Setup-2 and Setup-3).** Gray-highlighted cells are the weighted average of F1 scores in the same source and target languages retrieved from Table 3. The other cells represent the column-based recovery scores corresponding to the given source language. The best recovery ratios are given in bold for each target language. The recovery scores are computed specifically for the models, i.e., the denominator is the gray-highlighted cell in the column of a model. For instance, the F1 score is 0.879 when the source and target are English (see Table 3); also, when the source is Chinese and the target is English the F1 score is 0.519. Thus, the recovery ratio (Eq. 3) of Chinese→English is $\frac{0.519}{0.879} = 59\%$. The results are used to answer **RQ-1** and **RQ-2**.

Model	**Source/Target**	English	Chinese	Arabic	Turkish	Polish
mBERT	English	0.879	13%	17%	**82%**	38%
	Chinese	59%	0.991	20%	59%	48%
	Arabic	25%	75%	0.998	42%	**95%**
	Turkish	**80%**	68%	24%	0.894	40%
	Polish	40%	**80%**	**77%**	43%	0.771
XLM-R	English	0.758	16%	9%	**80%**	23%
	Chinese	77%	0.996	7%	68%	15%
	Arabic	25%	**79%**	0.997	30%	**92%**
	Turkish	**95%**	46%	10%	0.885	36%
	Polish	49%	78%	**77%**	43%	0.794
mDeBERTa	English	0.882	55%	39%	**90%**	49%
	Chinese	67%	0.996	12%	68%	20%
	Arabic	31%	**82%**	0.997	41%	**93%**
	Turkish	**90%**	70%	38%	0.901	59%
	Polish	38%	81%	**86%**	39%	0.787

language for a high-resource target language, e.g., English and Arabic in our study. For example, Polish can be used as a source language for Arabic, and Turkish can be used as a source language for English, (**RQ2**).

We conclude that multilingual Transformer-based models, e.g., mDeBERTa, performs well even if the source language is different from the target language. These promising results show that a multilingual model can be used for a low-resource language, although the target language is not available in terms of training resources.

9 Limitations and Ethical Consideration

In this section, we discuss the limitations and challenges encountered in our study, including the scarcity of non-English misinformation detection datasets, the binary labeling approach, and the difficulties associated with using microblog text from social media platforms.

9.1 Datasets

Due to the limited availability of non-English misinformation detection social media datasets, we had to combine multiple datasets focusing on different topics and collected at different time periods. This diversity in the datasets could potentially introduce bias into our research. Ideally, a multilingual dataset collected during the same time period and on the same topic would be preferable for observing the transfer ability between languages. However, due to the limitations of misinformation detection datasets in low-resource languages, we were unable to create such a setup.

9.2 Labels

The datasets we utilized have binary labels in terms of veracity. While this approach provides a simple and straightforward way to label data, it may oversimplify the complexity of misinformation and disinformation. Binary labels do not account for different levels of reliability and accuracy. Furthermore, they may fail to capture cultural and sociopolitical variations, thereby limiting the model's ability to generalize well to different contexts.

9.3 Usage of Microblog Text

Texts obtained from social media platforms can be noisy and contain a mixture of multiple languages within a single text. Additionally, the quality of these texts can be low. These factors can pose challenges to language transfer ability and can decrease the accuracy of misinformation/disinformation detection. Moreover, inherent biases present in social media platforms can also influence the model and introduce bias into its predictions.

9.4 Ethical Consideration and Possible Use Cases

This paper acknowledges and addresses several ethical considerations inherent in the research and development of fake news detection. Privacy and data protection are of utmost importance, and user data and personal information are treated with strict confidentiality throughout the research process. Moreover, we acknowledge broader societal impacts of misinformation detection such as the potential for censorship, and the effects of trust on social media.

We also anticipate that the experimental setup investigated throughout the paper can be used for other NLP problems. The transfer learning ability across multiple languages in other problems, e.g., rumor or stance detection and emotion recognition (Küçük and Can, 2020), need to be studied for further possibilities.

10 Conclusion

In order to observe cross-lingual few-shot transfer skills between languages, we carried out a number of experiments. For this purpose, multiple languages were used in a cross-lingual transfer learning structure employing multilingual pre-trained models. In this way, we provide a comparative examination of the performance of state-of-the-art methods for the misinformation detection task. We believe that this study will help future NLP researchers who plan to use the low-source language datasets in their cross-lingual study by giving them insight.

We observe that English can be used as a source language for the Turkish language depending on the dataset domain. Our most important observation is the context of the data is essential and we observe relatively better results for the transfer abilities between languages whose datasets are in the same domain. In future work, we will include other languages, such as Czech and Finnish, to observe the effects of agglutinative patterns of those languages between Turkish. We also plan to improve our study into several social media platforms, such as Facebook posts and Instagram content to investigate the effect of the social media domain on the datasets.

References

Firoj Alam, Fahim Dalvi, Shaden Shaar, Nadir Durrani, Hamdy Mubarak, Alex Nikolov, Giovanni Da San Martino, Ahmed Abdelali, Hassan Sajjad, Kareem Darwish, et al. 2021. Fighting the covid-19 infodemic in social media: a holistic perspective and a call to arms. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 913–922.

Muhammad Arif, Atnafu Lambebo Tonja, Iqra Ameer, Olga Kolesnikova, Alexander Gelbukh, Grigori Sidorov, and Abdul Gafar Manuel Meque. 2022. Cic at checkthat! 2022: multi-class and cross-lingual fake news detection. *Working Notes of CLEF*.

Ethem F. Can, Aysu Ezen-Can, and Fazli Can. 2018. Multilingual sentiment analysis: An rnn-based framework for limited data.

Mingxi Cheng, Songli Wang, Xiaofeng Yan, Tianqi Yang, Wenshuo Wang, Zehao Huang, Xiongye Xiao, Shahin Nazarian, and Paul Bogdan. 2021. A covid-19 rumor dataset. *Frontiers in Psychology*, 12.

Samuel Kai Wah Chu, Runbin Xie, and Yanshu Wang. 2021. Cross-language fake news detection. *Data and Information Management*, 5(1):100–109.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.

Limeng Cui and Dongwon Lee. 2020. Coaid: Covid-19 healthcare misinformation dataset. *arXiv preprint arXiv:2006.00885*.

Sajjad Dadkhah, Xichen Zhang, Alexander Gerald Weismann, Amir Firouzi, and Ali A. Ghorbani. 2023. TruthSeeker: The Largest Social Media Ground-Truth Dataset for Real/Fake Content.

Enyan Dai, Yiwei Sun, and Suhang Wang. 2020. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):853–862.

Arkadipta De, Dibyanayan Bandyopadhyay, Baban Gain, and Asif Ekbal. 2021. A transformer-based approach to multilingual fake news detection in low-resource languages. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 21(1).

Daryna Dementieva and Alexander Panchenko. 2021. Cross-lingual evidence improves monolingual fake news detection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop*, pages 310–320.

Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. SemEval-2017 task 8: RumourEval: Determining rumour veracity and support for rumours. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 69–76, Vancouver, Canada. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Arkin Dharawat, Ismini Lourentzou, Alex Morales, and ChengXiang Zhai. 2022. Drink bleach or do what now? covid-hera: A study of risk-informed health decision making in the presence of covid-19 misinformation. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1):1218–1227.

Jiangshu Du, Yingtong Dou, Congying Xia, Limeng Cui, Jing Ma, and S Yu Philip. 2021. Cross-lingual covid-19 fake news detection. In *2021 International Conference on Data Mining Workshops (ICDMW)*, pages 859–862. IEEE.

Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. 2019. Semeval-2019 task 7: Rumoureval 2019: Determining rumour veracity and support for rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 845–854, United States. Association for Computational Linguistics.

Mohamed Seghir Hadj Ameur and Hassina Aliane. 2021. Aracovid19-mfh: Arabic covid-19 multi-label fake news & hate speech detection dataset. *Procedia Computer Science*, 189:232–241.

Hicham Hammouchi and Mounir Ghogho. 2022. Evidence-aware multilingual fake news detection. *IEEE Access*, 10:116808–116818.

Fatima Haouari, Maram Hasanain, Reem Suwaileh, and Tamer Elsayed. 2020. Arcov19-rumors: Arabic covid-19 twitter dataset for misinformation detection. *arXiv preprint arXiv:2010.08768*.

Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2022. Few-shot cross-lingual stance detection with sentiment-based pre-training. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10):10729–10737.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

Stefan Helmstetter and Heiko Paulheim. 2018. Weakly supervised learning for fake news detection on twitter. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 274–277.

T. Hsu and S. A. Thompson. 2023. Disinformation researchers raise alarms about a.i. chatbots. (Accessed: 02-Apr-2023).

Andrzej Jarynowski. 2020. A dataset of media releases (Twitter, News and Comments, Youtube, Facebook) from Poland related to COVID-19 for open research.

Rohit Kumar Kaliyar, Anurag Goswami, and Pratik Narang. 2021. Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia tools and applications*, 80(8):11765–11788.

Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour verification. *arXiv preprint arXiv:1806.03713*.

Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. *ACM Comput. Surv.*, 53(1).

Yichuan Li, Bohan Jiang, Kai Shu, and Huan Liu. 2020. Mm-covid: A multilingual and multimodal data repository for combating covid-19 disinformation. *arXiv preprint arXiv:2011.04088*.

Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 3818–3824. AAAI Press.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 708–717, Vancouver, Canada. Association for Computational Linguistics.

Marion Meyers, Gerhard Weiss, and Gerasimos Spanakis. 2020. Fake news detection on twitter using propagation structures. In *Disinformation in Open Online Media: Second Multidisciplinary International Symposium, MISDOOM 2020, Leiden, The Netherlands, October 26–27, 2020, Proceedings 2*, pages 138–158. Springer.

Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. 2019. Fake news detection on social media using geometric deep learning. *arXiv preprint arXiv:1902.06673*.

Minh Nguyen and Zhou Yu. 2021. Improving named entity recognition in spoken dialog systems by context and speech pattern modeling. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 45–55, Singapore and Online. Association for Computational Linguistics.

Tam Nguyen, Matthias Weidlich, Bolong Zheng, Hongzhi Yin, Nguyen Hung, and Bela Stantic. 2019. From anomaly detection to rumour detection using data streams of social platforms. *Proceedings of the VLDB Endowment*, 12:1016–1029.

Jian-Yun Nie, Michel Simard, Pierre Isabelle, and Richard Durand. 1999. Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the web. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 74–81.

Parth Patwa, Shivam Sharma, Srinivas Pykl, Vineeth Guptha, Gitanjali Kumari, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. 2021. Fighting an infodemic: COVID-19 fake news dataset. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, pages 21–29, Cham. Springer International Publishing.

Shaina Raza and Chen Ding. 2022. Fake news detection based on news content and social contexts: a transformer-based approach. *International Journal of Data Science and Analytics*, 13(4):335–362.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*, 8(3):171–188.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017a. Fake news detection on social media: A data mining perspective. *SIGKDD Explor. Newsl.*, 19(1):22–36.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017b. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.

Weiwei Sun, Chuan Meng, Qi Meng, Zhaochun Ren, Pengjie Ren, Zhumin Chen, and Maarten de Rijke. 2021. Conversations powered by cross-lingual knowledge. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 1442–1451, New York, NY, USA. Association for Computing Machinery.

Cagri Toraman, Oguzhan Ozcelik, Furkan Şahinuç, and Fazli Can. 2022a. Not good times for lies: Misinformation detection on the russia-ukraine war, covid-19, and refugees. *arXiv preprint arXiv:2210.05401*.

Cagri Toraman, Furkan Şahinuç, and Eyup Yilmaz. 2022b. Large-scale hate speech detection with cross-domain transfer. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2215–2225.

Iulia Turc, Kenton Lee, Jacob Eisenstein, Ming-Wei Chang, and Kristina Toutanova. 2021. Revisiting the primacy of english in zero-shot cross-lingual transfer. *CoRR*, abs/2106.16171.

Vladimir Vapnik. 1999. *The nature of statistical learning theory*. Springer science & business media.

William Yang Wang. 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 (Volume 2: Short Papers)*, pages 422–426, Vancouver, Canada. Association for Computational Linguistics.

Weiming Wen, Songwen Su, and Zhou Yu. 2018. Cross-lingual cross-platform rumor verification pivoting on multimedia content. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3487–3496.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Jinxi Xu, Ralph Weischedel, and Chanh Nguyen. 2001. Evaluating a probabilistic model for cross-lingual information retrieval. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '01, page 105–110, New York, NY, USA. Association for Computing Machinery.

Chen Yang, Xinyi Zhou, and Reza Zafarani. 2021. Checked: Chinese covid-19 fake news dataset. *Social Network Analysis and Mining (SNAM)*.

Liu Zhuang, Lin Wayne, Shi Ya, and Zhao Jun. 2021. A robustly optimized BERT pre-training approach with post-training. In *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, pages 1218–1227, Huhhot, China. Chinese Information Processing Society of China.