2022 27th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran

Persian Stance Detection with Transfer Learning and Data Augmentation

1st Homa Nasiri
dept. Computer Engineering
Iran University of Science and Technology
Tehran, Iran
homa_nasiri@comp.iust.ac.ir

2nd Morteza Analoui
dept. Computer Engineering
Iran University of Science and Technology
Tehran, Iran
analoui@iust.ac.ir

2022 27th International Computer Conference, Computer Society of Iran (CSICC) | 978-1-6654-8027-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/CSICC55295.2022.9780479

***Abstract*— With increasing public access to social media, many dubious and inaccurate content is being generated and shared for profitable targets. This content is generated to attract audiences, increase revenue, impact people's decisions, and influence important events such as political elections. Manually, detecting this news is very time-consuming, costly, and tedious, so the automatic detection of this content has attracted the attention of many natural language processing researchers. In this research, we have investigated the effect of using EDA data augmentation methods and the ParsBERT pre-training model to solve the problem of scarcity and lack of annotated datasets in the Persian stance detection task. The results of this study indicate that we can identify the stance of news to the specific claim better than previous works with the help of data augmentation methods, content-based representation of the data, and the ParsBERT model.**

Keywords*— *Stance Detection, Transfer Learning, Data Augmentation, Fake News Detection, ParsBERT, EDA, BERT-based Model.

I. INTRODUCTION

In recent decades, extracting information from texts has attracted the attention of many researchers in various fields. With the spread of internet usage, the popularity of online tools, and the widespread availability of social networks such as Facebook, Twitter, and Instagram, lots of content have been accessed by analysts on various topics. These automatic analytics are used in different areas such as market trends, business analysis, product reviews, targeted advertising, providing a better user experience, elections, and referendums predictions, automated media monitoring, polls, and public health surveillance [1], [2]. Also, policymakers can better meet people's needs in political and public services by constantly monitoring people's opinions, criticisms, and demands [3].

This popularity and the ubiquitous availability of social media, along with controversial events such as Brexit, the presidential election, Coronavirus pandemic, Etc. in addition to the helpful and valuable content, has led to the outpouring of questionable content in traditional and online media [2], [4].

Fake news is produced for different purposes, such as attracting audiences, influencing people's opinions and decisions, increasing revenue through a click, and influencing important events such as political elections [4]. In addition, fake news about violent events in the real world can threaten public safety [5]. Thus, policymakers and the news industry face a significant challenge in detecting fake news, so automatic tools have become a basic necessity for detecting these kinds of news [2], [4].

Despite the availability of different tools, fake news detection is also a complex task for automated systems [6] and needs to be reduced to multi-steps. Gathering articles and understanding the reports of other news stations about the same subject would be a valuable start in fake news detection [4].

Stance can be defined as the expression of a speaker's point of view and judgment on a particular proposition [2], whether the author of a text favors the target or is opposed to it, or no inference is possible towards the target. Stance classification involves identifying the author's mental orientation towards a particular subject [7].

The main inputs of the claim-based stance detection model are the claim, which can be a rumor post or an article title, and the text, which can be the headline of the article or the body of the news.

One of the problems faced by stance detection is the lack of existing annotated datasets and the scarcity of some class entities; The dataset in the Persian language is no exception to this rule. As we all know, deep learning algorithms require vast amounts of data to perform better. Proper representation of this data is also essential to improve network performance.

Most of the previous research is on data gathering or network architecture and the type of features the network will use for classification. This study addressed scarcity and lack of data problems instead of focusing on network and types of features. With the help of data augmentation methods, an attempt was made to eliminate the data scarcity of some classes of this task by producing new samples in the train set. Data distribution in each class does not follow the same distribution due to the nature of the news in the real world. Most of the available news can contain information unrelated to the claim or do not provide information to the reader to refute or deny the claim, so the number of "*unrelated*" and "*discuss*" entities are much higher than the number of *"agree"* and *"disagree"* classes, and this would lead to the scarcity of *"agree"* and *"disagree"* classes in the dataset. The scarcity of these two classes causes the models not to identify their samples correctly, which can be solved by producing new samples of these two classes and adding some balance to the dataset.

With the data augmentation in minority class, we will continue to face the lack of entities in all classes, and only the unbalanced class distribution will improve slightly. Due to the lack of data, deep learning models will not have enough information to extract the patterns of each class entirely; To address this issue, the ParsBERT pre-trained model has shown that it can significantly help improve outcomes. This model was able to better identify the stance of each claim than the existing ones by combining pre-existing information, various structures, and knowledge which it had been seen in the past, with patterns extracted from stance detection data.

In addition to these cases, previous works typically used prediction-based embedding algorithms to detect the stance or rumor of the Persian language. By providing the exact representation for a word used in different content, these algorithms do not provide helpful information about the word's content to the model. In this research, we provide different representations for each word based on its content with ParsBERT embedding to the model to better identify each class's labels. Finally, with the help of the ASHA algorithm, suitable hyper-parameters were found and used for its training process.

The rest of the paper is structured as follows. Section II discusses previous works in the Stance detection task. Section III describes the proposed model. Section IV gives a summary of the dataset and the experiment details. Section V presents the results, and finally, section VI concludes the paper and future works.

II. Related Work

In general, Stance detection can be interpreted as a multi-class classification where a text is classified as *"agree"*, *"disagree"*, *"discuss"*, or *"unrelated"* toward a specific claim.

Most studies used traditional machine learning methods for classifying stances. They employ different algorithms such as SVM [8]–[13] Logistic Regression [9], [14], [15], naive base [1], Decision tree [1], [8], [16] and random forest [1], [9], [10], [17]. KNN [18] and k-means clustering [16] also has been used on this task. These methods work with different features. Most contributors take an entirely textual feature-oriented approach, while others incorporate information around the news spreading and propagation source [19].

Many stance detection studies also used deep neural networks such as LSTM [17], [20]–[25] and CNN [1], [26], [27] and their combinations. In [6], they used bidirectional RNNs, max-pooling, and neural attention mechanisms to represented news and combined these representations with external similarity features to address the problem.

In another work, [28] employed deep RNNs to calculate the neural embedding, n-gram weighted, and Bag of Words technique for statistical features and used greedy features engineering to extract handmade features.

Stacked layers were introduced as another approach too. [29] used stacked bi-LSTM, and [30] used stacked CNN as their based architecture. In [4], LSTM-CNN, and in [27], bidirectional GRU and CNN with attention mechanism are employed.

Also, in another study, a stance detection language model was proposed that used transfer learning with the Roberta model [31]. [32] presented a hidden language model by a BERT-based model. [33] were also able to improve the results by BERT-based embedder, XLNet, and Roberta.

Studies on different languages other than English got some attention too. Some works were done in Arabic, Danish, Ukrainian [2], Italian [3], Russian [14], Hindi [10], and recently in Persian. Most of the work done in Persian is in rumor detection and, there is no specific study on stance detection. [17] did the only work. They released the Persian stance detection dataset and used traditional methods and stacked-LSTM for evaluating their dataset.

III. Proposed Model

The proposed model uses transfer learning and data augmentation as based techniques to detect the stance of a text towards a claim. Fig.1 shows the architecture of the proposed model for stance detection, and its steps are as follows:

- *Data Augmentation:* Generating some samples with EDA algorithms.
- *Text Pre-processing:* Cleaning and converting the raw text into a list of words or tokens with ParsBERT tokenizer to utilize in the model.
- *Representation:* Extracting position embedding and word embeddings by ParsBERT.
- *Hyperparameter Finder:* Finding optimum hyperparameter by ASHA algorithms.
- *Network Architecture:* Adding a linear layer at the top of ParsBERT pre-trained model and applying a softmax function, and predicting the result of stance detection in four classes.

A. Data augmentation and Down Sampling

EDA (Easy Data Augmentation) consists of four simple but powerful operations: random deletion, random swap, random insertion, and synonym replacement [34]. We use these four techniques to generate augmented sentences for the minority class. Also, We down sample some of the majority class entities of article-claim datasets.

B. Text Pre-processing

As we know, The first step in natural language processing is data pre-processing, which helps to improve the quality of data and extract better meaning from it. In this paper, normalization operations have been performed on the existing texts. In this process, the random form of each person's writing becomes a standard that the machine can understand. Subsequently, sequences such as numbers, stop words, punctuations, extra spaces, and unwanted ASCII codes were cleared from the text.

C. Hyperparameter Finder

Successive halving (ASHA) [35] is an algorithm based on the multi-armed bandit methodology. The ASHA algorithm is a way to combine random search with principled early stopping in an asynchronous way.

We use the ASHA algorithm with minimizing loss function criteria to find the optimum hyperparameters (Learning Rate, Batch Size, and number of Epochs) for fine-tuning the model.

D. Network Architecture

We use ParsBERT [36], a monolingual BERT model for the Persian language, as our based network architecture and a linear layer at the top of that with a softmax function to determine the relevant class. ParsBERT is based on $BERT_{BASE}$ model architecture [37], which has a multi-layer bidirectional transformer. In particular, they use the original $BERT_{BASE}$ configuration: 12 hidden layers, 12 attention

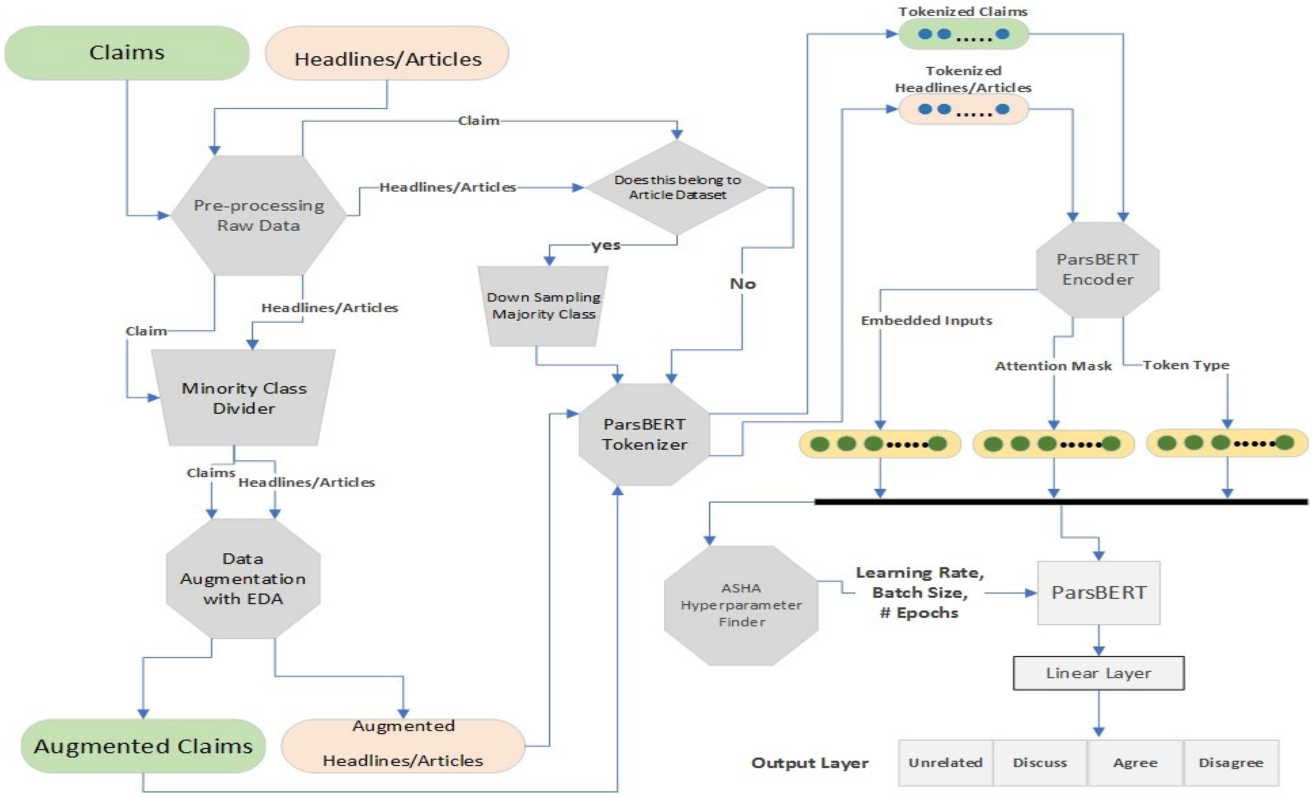

Fig. 1. General architecture of the proposal model.

heads, 768 hidden sizes. The total trainable parameters is 110M with this configuration [36]. We transfer pre-trained ParsBERT parameters and fine-tune them with our data.

For model optimization, AdamW optimizer with epsilon=1e-8 is used for 8 epochs. We set the batch size to 16 for headline-claim and 8 for article-claim. Each sequence in headline-claim contains 36 tokens and in article-claim contains 512 tokens. Finally, the linear scheduler learning rate is initiated to 7.19e-6 for headline-claim and 1.26e-6 for article-claim.

IV. Experiment

A. Dataset

We use the Persian stance classification dataset, which [17] gathered from *Shayeaat*[1] and *Fakenews*[2]. This dataset consists of two parts; the first part includes the claim with the news headline, and the next part is the claim with the article's body text. In particular, every article headline or body text received a label indicating whether the article is supporting (*"agree"* label), declining (*"disagree"* label), just reporting (*"discuss"* label), or irrelevant (*"unrelated"* label) to the claim [17]. The headline-claim includes 2029 entities, and the article-claim has 1997 entities. TABLE I illustrates the distribution of stance classes.

B. Implementation

The HuggingFace and the Pytorch framework have been used in Python programming language to implement this project. Tune Ray library also has been utilized as a hyperparameter optimization framework.

TABLE I.

Distribution of classes.

Type	Agree	Disagree	Discuss	Unrelated
headline-claim stance	405	164	802	658
article-claim stance	137	206	1068	586

C. Data Augmentation

We performed different experiments with EDA operations separately and combined these operations on the headline-claim dataset for selecting suitable augmentation operations. In the last experiment, the minority class was divided into four sections, and each operation was performed on each section. Then, we combined augmented sentences and added them to the train set. Model configurations in each experiment can be found in TABLE II.

D. ASHA

Since hyperparameter selection is one of the most important steps in model fine-tuning, the ASHA algorithm for batch size, learning rate, and the number of epochs is conducted as indicated in TABLE III.

V. Result

This section will present the result we got from different experiments. At first, we will discuss the experiment on Augmentation techniques, and after that, we will compare our model with previous work.

1 shayeaat.ir
2 fakenews.ir

TABLE II.

MODEL CONFIGURATIONS FOR EACH AUGMENTATION EXPERIMENTS.

Augmentation technique	Learning rate	Batch size	Epochs
Synonym replacement	4.0013e − 06	16	8
Random insertion	3.7142e − 06	16	8
Random swap	1.2153e − 06	4	10
Random deletion	3.7142e − 06	16	8
Mixed combination	4.6191e − 06	16	6

TABLE III.

SEARCH SPACE OVER CHOSEN HYPERPARAMETERS.

Hyperparameter	Considered Configurations
Learning rate	(1e-6, 1e-1)
Batch size	4;8;16
Epochs	8;16

A. Augmentation

As we mentioned in section IV-C, one of our experiments was on EDA data augmentation techniques. Those results can be found in TABLE IV. The results show that combining all techniques would be more suitable for this data.

TABLE IV.

DIFFERENT AUGMENTATION TECHNIQUES RESULTS.

Augmentation technique	Accuracy	F1
Synonym replacement	71.40	71.47
Random insertion	71.47	71.44
Random swap	72.45	72.41
Random deletion	71.8	71.69
Mixed combination	**73.77**	**73.73**

B. Comparison

We used 10-fold cross-validation to assess the generalizability, independence, and robustness of the model from data. Our model gained 74.64±1.64 on accuracy and 74.61±1.61 on F1 on headline-claim data. Their box and whisker plot of F1 on all results showed in Fig.2.

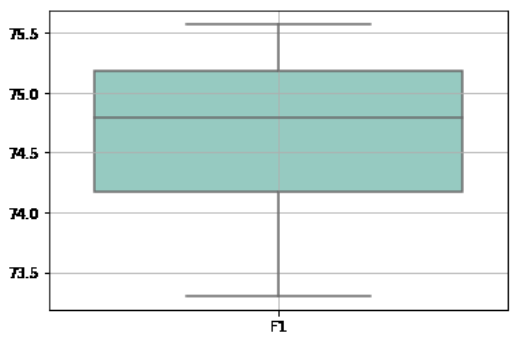

Fig. 2. The confdence interval of the proposed model based on F1.

The work of [17] is the only comparable work in Persian stance detection. We compare our model with their best model in TABLE V. As can be seen, transfer learning and data augmentation have a significant impact on improving the stance detection task.

This model was able to detect the stance of each claim better than the previous work by combining patterns extracted from stance detection data and pre-existing information and knowledge extracted from a large corpus, which ParsBERT was trained on in different fields. This information has come to aid with the augmentation of new samples from the minority class; it has compensated for the lack of quality data in this field. Also, the use of contextual-based embedding as ParsBERT provided a different representation of each word according to its content so that the model could better understand its inputs and has achieved significant improvement.

TABLE V.

COMPARISON OF THE PROPOSED MODEL WITH THE MODEL PRESENTED BY ZERHREN ET AL.

Model	Headline-Claim		Article-Claim	
	Accuracy	F1	Accuracy	F1
The proposed model	**74.64**	**74.61**	**76.33**	**75.64**
Zarharan et al. model	68.96	67.49	72	71

VI. CONCLUSION AND FUTURE WORKS

This study proposed a Persian stance detection model, based on the headline or body of the news, to identify the author's stance toward a specific claim. Contrary to existing works that emphasized the type of input features, this study sought to examine the impact of data augmentation and transfer learning on this task. The proposed model using the EDA data augmentation method and the ParsBERT pre-trained model with content-based embedding slightly improved the scarcity and lack of annotated data.

Future works can improve the proposed model in several ways. Since stance detection is closely related to sentiment analysis, the effectiveness of news polarity on stance detection can be studied. Some studies can generally examine the impact of different features with the proposed model by training another network with textual features and using it as part of ensemble learning to use both the advantages of pre-trained models and the advantages of textual features. In addition, we can use a CNN network or summarization methods to summarize the news content of the article-claim data, and the summarized content can be considered one of the inputs to the ParsBERT network. It is also possible to collect other raw data, add them to the dataset using clustering methods, and discuss its impact on the final results.

REFERENCES

[1] D. Küçük and F. Can, “Stance detection: A survey,” ACM Comput. Surv. CSUR, vol. 53, no. 1, pp. 1–37, 2020.

[2] M. Hardalov, A. Arora, P. Nakov, and I. Augenstein, “A survey on stance detection for mis-and disinformation identification,” ArXiv Prepr. ArXiv210300242, 2021.

[3] A. T. Cignarella, M. Lai, C. Bosco, V. Patti, and R. Paolo, “SardiStance@ EVALITA2020: Overview of the Task on Stance Detection in Italian Tweets,” in EVALITA 2020 Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, 2020, pp. 1–10.

[4] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, “Fake news stance detection using deep learning architecture (cnn-lstm),” IEEE Access, vol. 8, pp. 156695–156706, 2020.

[5] R. Oshikawa, J. Qian, and W. Y. Wang, “A survey on natural language processing for fake news detection,” ArXiv Prepr. ArXiv181100770, 2018.

[6] L. Borges, B. Martins, and P. Calado, “Combining similarity features and deep representation learning for stance detection in the context of checking fake news,” J. Data Inf. Qual. JDIQ, vol. 11, no. 3, pp. 1–26, 2019.

[7] P. Krejzl, “Stance detection and summarization in social networks,” 2018.

[8] A. Addawood, J. Schneider, and M. Bashir, “Stance classification of twitter debates: The encryption debate as a use case,” in Proceedings of the 8th International Conference on Social Media & Society, 2017, pp. 1–10.

[9] A. Tsakalidis, N. Aletras, A. I. Cristea, and M. Liakata, “Nowcasting the stance of social media users in a sudden vote: The case of the Greek Referendum,” in Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 367–376.

[10] S. Swami, A. Khandelwal, V. Singh, S. S. Akhtar, and M. Shrivastava, “An english-hindi code-mixed corpus: Stance annotation and baseline system,” ArXiv Prepr. ArXiv180511868, 2018.

[11] P. Sobhani, S. Mohammad, and S. Kiritchenko, “Detecting stance in tweets and analyzing its interaction with sentiment,” in Proceedings of the fifth joint conference on lexical and computational semantics, 2016, pp. 159–169.

[12] D. Küçük and F. Can, “Stance detection on tweets: An svm-based approach,” ArXiv Prepr. ArXiv180308910, 2018.

[13] K. Kucher, C. Paradis, M. Sahlgren, and A. Kerren, “Active learning and visual analytics for stance classification with ALVA,” ACM Trans. Interact. Intell. Syst. TiiS, vol. 7, no. 3, pp. 1–31, 2017.

[14] N. Lozhnikov, L. Derczynski, and M. Mazzara, “Stance prediction for russian: data and analysis,” in International Conference in Software Engineering for Defence Applications, 2018, pp. 176–186.

[15] K. Kucher, C. Paradis, and A. Kerren, “Visual Analysis of Sentiment and Stance in Social Media Texts.,” in EuroVis (Posters), 2018, pp. 49–51.

[16] V. Simaki, C. Paradis, and A. Kerren, “Stance classification in texts from blogs on the 2016 British referendum,” in International Conference on Speech and Computer, 2017, pp. 700–709.

[17] M. Zarharan, S. Ahangar, F. S. Rezvaninejad, M. L. Bidhendi, M. T.Pilevar, B. Minaei, and S. Eetemadi, “Persian Stance Classification DataSet.” inTTO, 2019.

[18] G. G. Shenoy, E. H. Dsouza, and S. Kübler, “Performing stance detection on Twitter data using computational linguistics techniques,” ArXiv Prepr. ArXiv170302019, 2017.

[19] M. Guderlei and M. Aßenmacher, “Evaluating Unsupervised Representation Learning for Detecting Stances of Fake News,” in Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 6339–6349.

[20] P. Wei, J. Lin, and W. Mao, “Multi-target stance detection via a dynamic memory-augmented network,” in The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018, pp. 1229–1232.

[21] Q. Sun, Z. Wang, Q. Zhu, and G. Zhou, “Stance detection with hierarchical attention network,” in Proceedings of the 27th international conference on computational linguistics, 2018, pp. 2399–2409.

[22] Q. Sun, Z. Wang, S. Li, Q. Zhu, and G. Zhou, “Stance detection via sentiment information and neural network model,” Front. Comput. Sci., vol. 13, no. 1, pp. 127–138, 2019.

[23] K. Dey, R. Shrivastava, and S. Kaushik, “Topical stance detection for Twitter: A two-phase LSTM model using attention,” in European Conference on Information Retrieval, 2018, pp. 529–536.

[24] B. Mavrin, “Statistical Modeling of Stance Detection,” 2017.

[25] G. Rajendran, B. Chitturi, and P. Poornachandran, “Stance-in-depth deep neural approach to stance classification,” Procedia Comput. Sci., vol. 132, pp. 1646–1653, 2018.

[26] S. Zhang, L. Qiu, F. Chen, W. Zhang, Y. Yu, and N. Elhadad, “We make choices we think are going to save us: Debate and stance identification for online breast cancer CAM discussions,” in Proceedings of the 26th International Conference on World Wide Web Companion, 2017, pp. 1073–1081.

[27] Y. Zhou, A. I. Cristea, and L. Shi, “Connecting targets to tweets: Semantic attention-based model for target-specific stance detection,” in International Conference on Web Information Systems Engineering, 2017, pp. 18–32.

[28] G. Bhatt, A. Sharma, S. Sharma, A. Nagpal, B. Raman, and A. Mittal, “Combining neural, statistical and external features for fake news stance identification,” in Companion Proceedings of the The Web Conference 2018, 2018, pp. 1353–1357.

[29] A. Hanselowski et al., “A Retrospective Analysis of the Fake News Challenge Stance-Detection Task,” in Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 1859–1874.

[30] M. Umer, S. Sadiq, M. Ahmad, S. Ullah, G. S. Choi, and A. Mehmood, “A novel stacked cnn for malarial parasite detection in thin blood smear images,” IEEE Access, vol. 8, pp. 93782–93792, 2020.

[31] C. Dulhanty, J. L. Deglint, I. B. Daya, and A. Wong, “Taking a stance on fake news: Towards automatic disinformation assessment via deep bidirectional transformer language models for stance detection,” ArXiv Prepr. ArXiv191111951, 2019.

[32] K. Kawintiranon and L. Singh, “Knowledge Enhanced Masked Language Model for Stance Detection,” in Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 4725–4735.

[33] G. Attardi and V. Slovikovskaya, “Transfer Learning from Transformers to Fake News Challenge Stance Detection (FNC-1) Task,” 2020.

[34] J. Wei and K. Zou, “EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks,” 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), 2019, pp. 6382–6388.

[35] L. Li et al., “A System for Massively Parallel Hyperparameter Tuning,” Proc. Mach. Learn. Syst., vol. 2, pp. 230–246, 2020.

[36] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, “Parsbert: Transformer-based model for persian language understanding,” Neural Process. Lett., pp. 1–17, 2021.

[37] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2019.