An Improvement in Transformer-Based Sentiment Analysis in Persian Twitter

Hadi Heydari (iD)
Faculty of Engineering
Kharazmi University
Tehran, Iran
h.heydari@khu.ac.ir

Ali Yazdani (iD)
Faculty of Computer Science
and Engineering
Shahid Beheshti University
Tehran, Iran
ali.yazdani@mail.sbu.ac.ir

Mojgan Farhoodi* (iD)
ICT Research Institute,
Tehran, Iran,
farhoodi@itrc.ac.ir

Abstract—**In the era of deep learning, transformer-based models have revolutionized natural language processing tasks, offering unparalleled performance in capturing contextual relationships. This paper delves into the realm of sentiment analysis in Persian Twitter, employing state-of-the-art transformer architectures. Through rigorous experimentation on a dedicated Persian sentiment dataset, we explore the capabilities of transformers in deciphering nuanced emotions expressed in tweets. The results demonstrate the potency of these models, highlighting their effectiveness in understanding the intricacies of sentiment within the Persian language. This study not only contributes insights into sentiment analysis but also underscores the transformative impact of transformer architectures in unlocking the expressive dynamics of Persian social media discourse. We trained multiple deep learning architectures based on transformers for sentiment analysis on Persian Twitter data, and in the test section, we achieved a 60.37% F-score.**

Keywords*: Sentiment Analysis, Persian Language, Deep Learning, Transformers, Social Media Sentiment Analysis.*
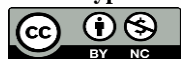
I. INTRODUCTION

Sentiment analysis, a pivotal facet of natural language processing, entails the computational exploration of textual data to discern the expressed sentiment, typically characterized as positive, negative, or neutral. As the proliferation of user-generated content on digital platforms continues, sentiment analysis plays a crucial role in deciphering public opinion, consumer feedback, and social sentiment. This process involves leveraging various techniques, ranging from traditional machine learning algorithms to sophisticated deep learning models, to accurately classify and quantify sentiments within a given text. Key methodologies encompass lexicon-based approaches, machine learning classifiers, and state-of-the-art neural networks, with advancements like BERT and its derivatives providing substantial improvements in sentiment analysis accuracy[1], [2].

Distinguishing sentiment from emotion is essential in understanding the nuanced nature of human expression. While sentiment refers to the overall positive, negative, or neutral orientation of a text,

---
* Corresponding Author

emotion delves deeper into the specific emotional states conveyed within the content. Sentiment analysis primarily focuses on polarity and subjective judgment, discerning the overarching attitude or opinion expressed. On the other hand, emotion analysis aims to identify and categorize specific emotions such as joy, anger, sadness, or surprise embedded in the text. Emotion analysis often involves a more granular examination of linguistic cues, facial expressions, and contextual elements to accurately infer the diverse spectrum of human emotions within textual data[3], [4]. Recognizing these distinctions is paramount for refining the accuracy and depth of sentiment analysis models and advancing the comprehension of human communication in the digital realm.

Sentiment analysis is a multidisciplinary science that seeks to understand and measure emotions and opinions using various tools and techniques[5]. In recent years, with technological advancements and the increased use of social networks, sentiment analysis has become a powerful tool for a deeper understanding of collective beliefs and social trends[6]. In the active and dynamic environment of Twitter, sentiment analysis serves as a key tool for the critique and examination of users' messages and opinions in real-time. By employing algorithms and deep learning models, sentiment analysis on Twitter enables the identification of satisfaction levels, displeasure, or even the impact of a specific topic on the community[7].

Some applications of sentiment analysis on Twitter include various substructures, such as:

- **Social Research:** Analyzing users' opinions and sentiments as a crucial source for understanding social behaviors and tendencies.

- **Brand Monitoring:** Predicting and measuring the positive or negative effects of crises, advertisements, or significant events on brand perceptions.

- **Customer Support:** Quickly tracking feedback and reviews for the rapid improvement of services and products.

- **Market Trend Prediction:** Early detection of changes in public opinions that can have a significant impact on business decisions[1], [5], [6].

By analyzing sentiment on Twitter, one can act as a keen observer of the dynamics and trends of modern society, using this information for process improvement and intelligent decision-making[7].

Deep learning has revolutionized the field of sentiment analysis by providing advanced tools for automatic feature extraction and representation learning. Unlike traditional machine learning approaches that rely on handcrafted features, deep learning models, particularly neural networks, can automatically learn intricate patterns and representations from raw data[2]. In summary, the application of deep learning in sentiment analysis has significantly enhanced the accuracy and efficiency of emotion and opinion extraction from textual data. The ability of neural networks to autonomously learn

relevant features and understand complex contextual nuances has propelled sentiment analysis into a new era of sophistication and effectiveness[2]. Our goal in this article is to train a deep learning model on available datasets in the Persian language, aiming to advance sentiment analysis capabilities and contribute to the development of sophisticated natural language processing tools in the Persian-speaking context.

Continuing, Section 2 delves into the comprehensive exploration of available Persian language datasets for sentiment analysis, followed by Section 3, which details the architecture and design considerations for implementing a robust deep learning model. In Section 4, we present the training process and fine-tuning strategies applied to optimize the model's performance on Persian sentiment analysis tasks. Finally, Section 5 discusses the results and evaluates the effectiveness of the proposed deep learning model, providing insights into its potential applications and implications for future research in sentiment analysis in the Persian language.

II.   RELATED WORKS

Sentiment analysis in the Persian language has gained significant attention in recent years, reflecting a growing interest in understanding the sentiments expressed in this linguistic domain. Previous research has laid the essential groundwork in various aspects of sentiment analysis, addressing both challenges and opportunities unique to the Persian language[5], [7]–[9].

BERT (Bidirectional Encoder Representations from Transformers) revolutionized natural language processing with its bidirectional context-aware model, capturing nuanced linguistic relationships. Devlin et al.'s seminal paper in 2018 laid the groundwork for this transformative approach [10]. Building upon BERT's success, XLM-RoBERTa in 2019, introduced by Conneau et al. in [11], extended the model to cross-lingual pre-training, enabling effective knowledge transfer across multiple languages. XLM-RoBERTa demonstrated state-of-the-art results on diverse cross-lingual benchmarks. Tailoring this paradigm to Persian, Mollanorozy et al. in [12], introduced ParsBERT in 2023, a model incorporating BERT's architecture with linguistic adaptations specific to Persian, including orthographic and morphological considerations. ParsBERT showcased promising results in cross-lingual transfer learning with Persian, contributing significantly to Persian language processing within the BERT framework.

Birjali et al. in [1], provide a comprehensive survey on sentiment analysis, offering insights into different methodologies and challenges. Wankhade et al. in [6], further expand on sentiment analysis methods, applications, and challenges, providing a broader perspective on the field. These foundational surveys set the stage for understanding the landscape of sentiment analysis across languages. Xu et al. in [5], delve specifically into social media-based sentiment analysis, highlighting emerging trends and challenges. This is particularly relevant in the context of Persian sentiment analysis, given the prevalence of social media platforms in the region. Rajabi et al. in [13] and

Nazarizadeh et al. in [14], present surveys on sentiment analysis methods in Persian text, offering a detailed exploration of the existing approaches and datasets. This foundational understanding is crucial for the development and improvement of sentiment analysis models tailored to the Persian language.

The work by Mollanorozy et al. in [15], introduces the concept of cross-lingual transfer learning with Persian, showcasing the potential for leveraging knowledge from other languages in sentiment analysis tasks. The study explores the effectiveness of cross-lingual transfer learning for Part-of-Speech tagging using Persian as a focal language, revealing its potential as a beneficial resource for diverse low-resource languages, with task-dependent variations in language similarity benefits observed across POS tagging and sentiment analysis tasks. They achieved a 0.8585% of F1 with fine-tuning XLM-RoBERTa model.

Ghasemi et al. in [8], and the DeepSentiParsBERT model in [9], focus on deep learning approaches for Persian sentiment analysis, emphasizing the importance of advanced neural network architectures. The study in [8], introduces a cross-lingual deep learning framework for sentiment analysis in Persian, leveraging English training data through cross-lingual embeddings, demonstrating superior performance compared to state-of-the-art monolingual techniques, with a notable 22% improvement in static embedding and 9% in dynamic embedding, and emphasizing the model's generalizability to any low-resource language when cross-lingual embeddings are available for the source–target language pair. In [9], they achieved 91.57% of F1-Score on the Digikala corpus.

Dehghani and Yazdanparast in [7], contribute to the field by applying the ParsBERT embedding model with Convolutional Neural Network (CNN) to analyze sentiment in Persian political tweets, showcasing the practical applications of sentiment analysis in real-world scenarios. The paper introduces a novel sentiment analysis methodology for Persian political tweets, utilizing various encoding methods and machine learning techniques, with CNN+BiLSTM using ParsBERT embeddings demonstrating higher robustness, achieving a score of 0.89 and 0.71 on two datasets of Persian political tweets, marking a significant advancement in the analysis of this specific subfield of Persian tweets.

The comparative studies by Dashtipour et al. in [16] and the subsequent hybrid frameworks explore different feature combinations and context-aware multimodal approaches for Persian sentiment analysis [17], [18]. These studies provide valuable insights into the effectiveness of various methodologies. In [16], they achieved 74.85% of Accuracy by using unigram + bigram. The work in [17], introduces a hybrid framework for concept-level sentiment analysis in Persian, combining linguistic rules and deep learning, outperforming state-of-the-art approaches and DNN classifiers with margins of 10–15% and 3–4%, respectively, on benchmark Persian product and hotel

reviews corpora. In [17], they introduced a first-of-its-kind Persian multimodal dataset and a context-aware multimodal sentiment analysis framework, achieving better performance (91.39%) through the contextual integration of textual, acoustic, and visual features compared to unimodal features (89.24%).

Also, the ensemble-based classification approach proposed by Dashtipour et al. in [19], and the application of deep learning in Persian movie reviews in [20]. The paper presents an ensemble classifier for Persian sentiment analysis, combining shallow and deep learning algorithms, achieving an impressive accuracy rate of up to 79.68% and showcasing improvements over state-of-the-art approaches. Dastgheib et al. in [20], further highlight the diverse range of methods employed in Persian sentiment analysis. They introduced a hybrid method, combining structural correspondence learning (SCL) and convolutional neural network (CNN), to improve sentiment classification for Persian text data, demonstrating that the proposed SCL+CNN hybrid method enhances classification results for two domains by more than 10 percent.

III. CONTRIBUTION

In this Section we explain our used datasets for train and testing, and then we define a new classifier architecture for text classification with Large Language Models (LLMs). Given the limitations of available open-source models for the Persian language, our approach involved leveraging open-source multilingual models to address the task at hand. Remarkably, our investigation yielded superior results compared to those obtained through the utilization of dedicated Persian models. The findings presented here in the significance of a multilingual model in overcoming the challenges posed by the scarcity of open-source datasets and Persian LLMs.

A. Dataset

We used an open-source dataset on GitHub[1] that is extracted from Twitter and sentiment labelled with five labels[21]. This label consists of low positive, positive, very positive, neutral, and negative so we changed this label to 3 labels. As can be seen in Table 1, this data is very imbalanced. For the test dataset, we used of dataset of the fourth step of 1st ParsiAzma[2] Challenge in sentiment analysis.

TABLE I. COMPARISON OF THE LABEL COUNTS

Label	Count
Positive	1619
Neutral	311
Negative	3070

B. Models

We defined a new architecture classifier for LLMs, in this paper that has better accuracy in text classification tasks. As previously alluded to, the extant landscape of pre-trained multilingual models for the Persian language does not have enough accuracy.

---

[1] https://github.com/behzadbb/Persian-Sentiment-Analysis-Text-Classification

[2] https://parsiazma.ir/

The outcomes revealed a discernible enhancement in accuracy when employing multilingual language models, exemplified by the noteworthy performance of XLM-RoBERTa (Unsupervised Cross-lingual Representation Learning at Scale, which is a scaled cross-lingual sentence encoder). These findings contribute empirical evidence substantiating the efficacy of multilingual language models in the context of Persian language Classification tasks. As previously alluded to, the extant landscape of pre-trained multilingual models for the Persian language does not have enough accuracy. The outcomes revealed a discernible enhancement in accuracy when employing multilingual language models, exemplified by the noteworthy performance of XLM-RoBERTa (Unsupervised Cross-lingual Representation Learning at Scale, which is a scaled cross-lingual sentence encoder). These findings contribute empirical evidence substantiating the efficacy of multilingual language models in the context of Persian language Classification tasks.

As shown in figure 1, The architectural framework devised for the classification task involving XLM-Roberta is predicated on an Encoder/Decoder
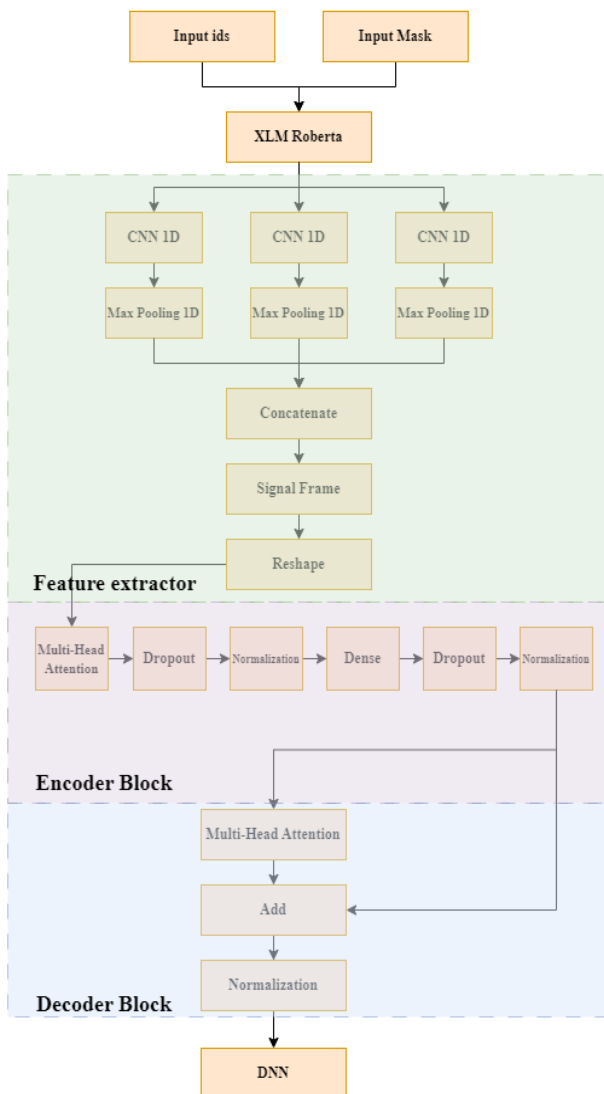
Fig. 2.   Model Architecture

paradigm. Initially, a feature extraction process is instituted, emanating from the last hidden state of XLM-Roberta's Language Model (LLMs). This extracted tensor undergoes a tripartite traversal through parallel 1-dimensional Convolutional Neural Networks (CNNs), with the resultant outputs from each CNN serving as inputs to subsequent Max-Pooling operations. The ensuing outputs of these Max-Pooling layers are amalgamated and channelled into a signal frame layer, followed by a reshaping operation to render it amenable for integration into the Encoder block.

Within this architectural framework, each convolutional layer operates distinctively due to the inclusion of parallel filters of varying sizes. This design choice enables the extraction of features at
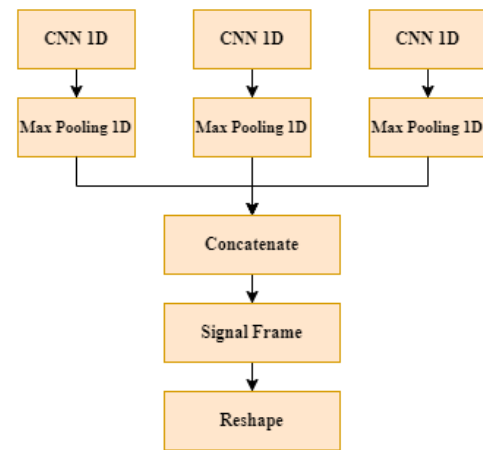
Fig. 1.   Feature Extraction

multiple levels, thereby ensuring that the derived features encapsulate pertinent information conducive to sentiment analysis. Moreover, this architectural configuration has demonstrated efficacy beyond the domain of sentiment analysis, exhibiting notable effectiveness in tasks such as text emotion recognition[22]. Figure 2 shows this part of the architecture.

Within the Encoder block, a confluence of multi-head attention mechanisms, dropout layers, normalization, and dense layers is orchestrated. The incorporation of these elements contributes to the hierarchical representation of features. Concurrently, the Decoder block is instantiated, wherein the outputs of the Encoder layers are subjected to a multi-head attention mechanism. The resulting outputs are augmented with the outputs of the Encoder layers and subsequently normalized. This composite output is then fed into a deep neural network to effectuate predictions. Figures 3 and 4 show the encoder and decoder blocks of the architecture, respectively.

The delineation of the Encoder/Decoder architecture not only encapsulates the salient components but also provides a comprehensive understanding of the orchestrated interplay of various layers, mechanisms, and transformations in the pursuit of robust classification within the XLM-Roberta framework.
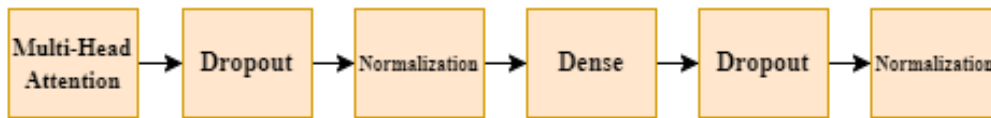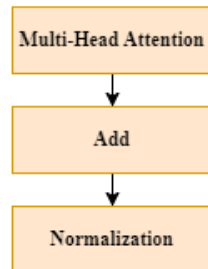
Fig. 4.   Encoder Block

Fig. 3.   Decoder Block

TABLE II.   RESULTS

Model	Macro F1
XLM-Roberta	44.59%
Perspolix	47.23%
XLM-Roberta + Feature extractor	57.65%
Perspolix + Feature extractor + Encoder/Decoder	58.75%
XLM-Roberta + Feature extractor + Encoder/Decoder	**60.37%**

C. Training Phase

During the training phase, we configured the optimization process utilizing the Adam optimizer, with a specified learning rate of 3e-5. To regulate sequence lengths, a maximum length constraint of 64 tokens was imposed, and training instances were processed in batches of size 4. The training regimen spanned 30 epochs, enabling the model to iteratively learn from the dataset.

In order to mitigate the risk of overfitting, a prudent approach was adopted through the incorporation of cross-validation. This technique facilitated the partitioning of the dataset into subsets, ensuring that the model's performance was not contingent solely on a particular training-validation split. We utilized 80% of the data for model training and reserved the remaining 20% for validation purposes.

IV.   EXPERIMENTS

In this section, we delineate the outcomes derived from our assessment of models characterized by identical properties and trained on the same dataset. The results of the model evaluations are presented in Table 2, providing a comprehensive overview of our findings.

We employed two pre-trained models, namely the Multilingual Twitter Sentiment Analysis model developed by Cardiff and Perspolix (Persian Political Tweet XLM-Roberta Large) [23], [24]. Subsequently,

all models underwent fine-tuning on the initial datasets, followed by rigorous testing on the designated test dataset. As observed, the integration of Feature Extractor and Encoder/Decoder blocks resulted in a significant augmentation of the F1-Score Macro outcomes.

V.   CONCLUSION

In conclusion, our investigation in this study has elucidated that the utilization of a composite approach involving XLM-Roberta and a meticulously devised neural network architecture can yield a potent synergy for classification tasks. This endeavor not only contributes to the current state of research but also sets the stage for prospective initiatives, suggesting that the integration of diverse models and methodologies, encompassing multilingual models, could serve as a valuable arsenal in addressing persisting challenges related to dataset and model scarcity.

However, it is imperative to underscore the exigency for the expansion of open-source large language models and datasets tailored to the Persian language. This expansion plays a pivotal role in perpetuating the progressive evolution of natural language processing tasks in Persian, thereby cultivating a more intricate, comprehensive, and robust terrain for both research and practical application.

REFERENCES

[1]   M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," Knowl.-Based Syst., vol. 226, p. 107134, Aug. 2021, doi: 10.1016/j.knosys.2021.107134.

[2]   A. Yadav and D. K. Vishwakarma, "Sentiment analysis using deep learning architectures: a review," Artif. Intell. Rev., vol. 53, no. 6, pp. 4335–4385, Aug. 2020, doi: 10.1007/s10462-019-09794-5.

[3]   "New Avenues in Opinion Mining and Sentiment Analysis | IEEE Journals & Magazine | IEEE Xplore." Accessed: Dec. 06, 2023.   [Online].   Available: https://ieeexplore.ieee.org/abstract/document/6468032

[4]   S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, and L.-P. Morency, "Context-Dependent Sentiment Analysis in User-Generated Videos," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), R. Barzilay and M.-Y. Kan, Eds., Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 873–883. doi: 10.18653/v1/P17-1081.

[5]   Q. A. Xu, V. Chang, and C. Jayne, "A systematic review of social media-based sentiment analysis: Emerging trends and challenges," Decis. Anal. J., vol. 3, p. 100073, Jun. 2022, doi: 10.1016/j.dajour.2022.100073.

[6]   M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges,"

Artif. Intell. Rev., vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.

[7] M. Dehghani and Z. Yazdanparast, "Sentiment Analysis of Persian Political Tweets Using ParsBERT Embedding Model with Convolutional Neural Network," in 2023 9th International Conference on Web Research (ICWR), May 2023, pp. 20–25. doi: 10.1109/ICWR57742.2023.10139063.

[8] R. Ghasemi, S. A. Ashrafi Asli, and S. Momtazi, "Deep Persian sentiment analysis: Cross-lingual training for low-resource languages," J. Inf. Sci., vol. 48, no. 4, pp. 449–462, Aug. 2022, doi: 10.1177/0165551520962781.

[9] "DeepSentiParsBERT: A Deep Learning Model for Persian Sentiment Analysis Using ParsBERT | IEEE Conference Publication | IEEE Xplore." Accessed: Oct. 31, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10105414

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv, May 24, 2019. doi: 10.48550/arXiv.1810.04805.

[11] A. Conneau et al., "Unsupervised Cross-lingual Representation Learning at Scale." arXiv, Apr. 07, 2020. doi: 10.48550/arXiv.1911.02116.

[12] "ParsBERT: Transformer-based Model for Persian Language Understanding | Neural Processing Letters." Accessed: Dec. 06, 2023. [Online]. Available: https://link.springer.com/article/10.1007/s11063-021-10528-4

[13] Z. Rajabi and M. Valavi, "A Survey on Sentiment Analysis in Persian: a Comprehensive System Perspective Covering Challenges and Advances in Resources and Methods," Cogn. Comput., vol. 13, no. 4, pp. 882–902, Jul. 2021, doi: 10.1007/s12559-021-09886-x.

[14] A. Nazarizadeh, T. Banirostam, and M. Sayyadpour, "Sentiment Analysis of Persian Language: Review of Algorithms, Approaches and Datasets." arXiv, Nov. 09, 2022. doi: 10.48550/arXiv.2212.06041.

[15] S. Mollanorozy, M. Tanti, and M. Nissim, "Cross-lingual Transfer Learning with Persian," in Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP, L. Beinborn, K. Goswami, S. Muradoğlu, A. Sorokin, R. Kumar, A. Shcherbakov, E. M. Ponti, R. Cotterell, and E. Vylomova, Eds., Dubrovnik, Croatia: Association for Computational Linguistics, May 2023, pp. 89–95. doi: 10.18653/v1/2023.sigtyp-1.9.

[16] K. Dashtipour, M. Gogate, A. Adeel, A. Hussain, A. Alqarafi, and T. Durrani, "A Comparative Study of Persian Sentiment Analysis Based on Different Feature Combinations," in Communications, Signal Processing, and Systems, Q. Liang, J. Mu, M. Jia, W. Wang, X. Feng, and B. Zhang, Eds., in Lecture Notes in Electrical Engineering. Singapore: Springer, 2019, pp. 2288–2294. doi: 10.1007/978-981-10-6571-2_279.

[17] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, and A. Hussain, "A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks," Neurocomputing, vol. 380, pp. 1–10, Mar. 2020, doi: 10.1016/j.neucom.2019.10.009.

[18] K. Dashtipour, M. Gogate, E. Cambria, and A. Hussain, "A novel context-aware multimodal framework for persian sentiment analysis," Neurocomputing, vol. 457, pp. 377–388, Oct. 2021, doi: 10.1016/j.neucom.2021.02.020.

[19] K. Dashtipour, C. Ieracitano, F. Carlo Morabito, A. Raza, and A. Hussain, "An Ensemble Based Classification Approach for Persian Sentiment Analysis," in Progresses in Artificial Intelligence and Neural Systems, A. Esposito, M. Faundez-Zanuy, F. C. Morabito, and E. Pasero, Eds., in Smart Innovation, Systems and Technologies. , Singapore: Springer, 2021, pp. 207–215. doi: 10.1007/978-981-15-5093-5_20.

[20] "Entropy | Free Full-Text | Sentiment Analysis of Persian Movie Reviews Using Deep Learning." Accessed: Dec. 05, 2023. [Online]. Available: https://www.mdpi.com/1099-4300/23/5/596

[21] "GitHub - behzadbb/Persian-Sentiment-Analysis-Text-Classification: Persian Sentiment Analysis Text Classification." Accessed: Dec. 06, 2023. [Online]. Available: https://github.com/behzadbb/Persian-Sentiment-Analysis-Text-Classification

[22] A. Yazdani and Y. Shekofteh, "A Persian ASR-based SER: Modification of Sharif Emotional Speech Database and Investigation of Persian Text Corpora." arXiv, Nov. 18, 2022. doi: 10.48550/arXiv.2211.09956.

[23] "cardiffnlp/tweet_sentiment_multilingual · Datasets at Hugging Face." Accessed: Dec. 06, 2023. [Online]. Available: https://huggingface.co/datasets/cardiffnlp/tweet_sentiment_multilingual

[24] "StateOfTheArtAUT/perspolix-persian-political-tweet-xlm-roberta-large · Hugging Face." Accessed: Dec. 06, 2023. [Online]. Available: https://huggingface.co/StateOfTheArtAUT/perspolix-persian-political-tweet-xlm-roberta-large

Hadi Heydari received his B.Sc. degree in Industrial Engineering from Shahrood University of Technology, Semnan, Iran, in 2020. He is currently a Master's degree student in Industrial Engineering at Kharazmi University, Tehran, Iran. His current research interests include Artificial Intelligence, Data-Driven Robust Optimization and Decision-Making under Uncertainty.

Ali Yazdani received his B.Sc. degree in Computer Engineering from Razi University, Kermanshah, Iran, in 2020. He received his M.Sc. degree in Computer Science and Engineering at Shahid Beheshti University, Tehran, Iran. His current research interests include Artificial Intelligence, Automatic Speech Recognition and Speech Emotion Recognition.

Mojgan Farhoodi received her B.Sc. degree in Software Engineering and her M.Sc. degree in IT Engineering from Amirkabir University of Technology. She has been working as a researcher at the ICT Research Institute since 2010. Currently, he is working as a faculty member of the research institute in the fields of Information Retrieval, Natural Language Processing and Artificial Intelligence.