Original

Improving Persian Sentiment Analysis Results in Healthcare using a Hybrid BERT_TF-IDF Approach

Faezeh Forootan (PhD Candidate)[1], Seyed Raouf Khayami (PhD)[1]*, Pirooz Shamsinejad (PhD)[1], Reza Javidan (PhD)[1]

[1]Department of Computer, Engineering and Information Technology, Shiraz University of Technology, Shiraz, Iran

ABSTRACT

Background: With the rapid growth of patient-generated textual reviews, sentiment analysis has become essential for extracting meaningful insights in the healthcare sector. However, embedding models, which are typically pre-trained on general-domain corpora, struggle to capture domain-specific terminology and contextual nuances in healthcare content accurately. This limitation has led to increasing interest in hybrid embedding strategies that combine the strengths of multiple representation methods.

Objective: This study aims to improve the performance of Persian healthcare sentiment analysis by proposing a hybrid embedding approach capable of more accurately identifying domain-specific expressions and linguistic ambiguities.

Material and Methods: In this analytical study, a hybrid embedding model integrating Bidirectional Encoder Representations from Transformers (BERT) and Frequency–Inverse Document Frequency (TF-IDF) representations were developed and applied to a dataset of Persian healthcare-related comments. Then, the Bidirectional Gated Recurrent Unit (Bi GRU) algorithm was employed for sentiment classification, and its performance was subsequently compared with that of other machine learning and deep learning models.

Results: The proposed hybrid method achieved an accuracy of 83.05%. In comparison, models using only BERT and TF-IDF embeddings achieved accuracies of 63.46% and 80%, respectively.

Conclusion: The results highlight the superiority of the proposed hybrid embedding approach in capturing domain-specific vocabulary and resolving ambiguous expressions within healthcare reviews. The strong performance of the Bi-GRU classifier further demonstrates the importance of modeling semantic and long-term dependencies to enhance sentiment analysis accuracy in the healthcare domain.

Keywords

Data Mining; Sentiment Analysis; Medical; Machine Learning; Deep Learning

Introduction

Recent statistics indicate that approximately 5% of all Google searches are related to healthcare topics, and nearly 90% of users trust online medical information [1, 2]. Approximately 94% of patients seek healthcare data online, and 75% consider user comments highly influential due to their perceived authenticity [3]. Consequently, sentiment analysis has become an essential tool in healthcare decision making, as approximately 60% of medical professionals believe it enhances patient insights and the quality of communication [4, 5].

*Corresponding author: Seyed Raouf Khayami
Department of Computer, Engineering and Information Technology, Shiraz University of Technology, Shiraz, Iran
E-mail: khayami@sutech.ac.ir

Received: 29 October 2025
Accepted: 8 February 2026

Accordingly, researchers increasingly integrated sentiment analysis approaches to analyze patient experiences and optimize service delivery. Despite this progress, sentiment analysis in healthcare faces significant challenges due to the presence of specialized terminology, complex linguistic structures, and semantic ambiguity in patients' opinions. These difficulties are most evident in the word embedding stage. Since most embedding models are pre-trained on general corpora, such as Wikipedia, they often fail to interpret medical terminology accurately and overlook important contextual nuances [6-9]. Such limitations highlight the need for domain-specific or hybrid embedding approaches tailored to healthcare contexts.

In response, recent studies have emphasized hybrid frameworks that combine syntactic and semantic embeddings to improve text comprehension [10-13]. Although these models have demonstrated strong performance on English and Arabic healthcare datasets [11-13], their application to Persian particularly in healthcare sentiment analysis remains limited due to the language's rich morphological structure and the scarcity of annotated datasets [11].

The limited number of studies on Persian healthcare sentiment analysis dates back to the work of BokaeeNezhad and Deihimi [14], who achieved 85% accuracy on Persian COVID-related sentiment analysis using a word embedding plus Convolutional Neural Network (CNN) approach. Later, Taghizadeh et al. [15] introduced the Sina Bidirectional Encoder Representations from Transformers or SINA-BERT model (a domain specific pre trained BERT model for Persian medical and healthcare texts) and demonstrated that it outperforms Pars-BERT for sentiment analysis in Persian healthcare. To bridge this gap, the present analytical computational study develops a hybrid embedding model that integrates Term Frequency Inverse Document Frequency (TF-IDF) and Bidirectional Encoder Representations from Transformers (BERT). The combined embeddings are subsequently fed into a Bidirectional Gated Recurrent Unit (Bi GRU) classifier for sentiment prediction. Comparative experiments involving BERT only, TF IDF only, Bidirectional Long Short Term Memory (Bi LSTM), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost) models confirm the robustness and effectiveness of the proposed approach. To the best of our knowledge, this is the first study to employ a hybrid integration of TF-IDF and BERT combined with a Bi-GRU framework for Persian healthcare sentiment analysis. The results are expected to advance computational techniques for healthcare opinion mining and deepen the understanding of patient perspectives within Persian medical contexts.

Material and Methods

In this analytical study, the framework (Figure 1) combines TF IDF and BERT embeddings to capture structural and semantic features, thereby improving embedding efficiency and boosting the accuracy of Persian healthcare sentiment analysis.

Dataset

Although Persian constitutes approximately 1.3% of all web content, the scarcity of large and reliable Persian datasets continues to pose a major obstacle to the development of accurate sentiment analysis models, particularly in the healthcare domain [16, 17]. To address this limitation, a web crawling approach was employed to collect healthcare related user comments from the Nobat.ir website. The extracted texts were compiled into the dataset used in this study, consisting of 5,000 Persian healthcare comments. Initially, the comments lacked sentiment polarity labels; therefore, polarity was assigned based on their corresponding star ratings. The resulting dataset is imbalanced, comprising 2,134 positive, 1,208 neutral, and 1,658 negative comments. The length of the comments varied substantially in terms of the number of lexical tokens, as summarized in Table 1. These characteristics reflect the dataset's wide expressive range and highlight the complexity of user sentiment in Persian healthcare related comments.

Pre-processing

Pre-processing steps in sentiment analysis vary depending on the study objectives, linguistic features, and language-specific complexities [12]. The Persian language presents unique challenges, such as the coexistence of Arabic characters, extensive colloquial usage, inconsistent punctuation, and the absence of clear phrase boundaries. To ensure reliable and consistent results in analyzing Persian healthcare comments, several techniques were utilized:

Normalization converts unstructured text into a standardized form by unifying characters, converting numbers into words, and removing punctuation marks [18].

Tokenization is the process of segmenting

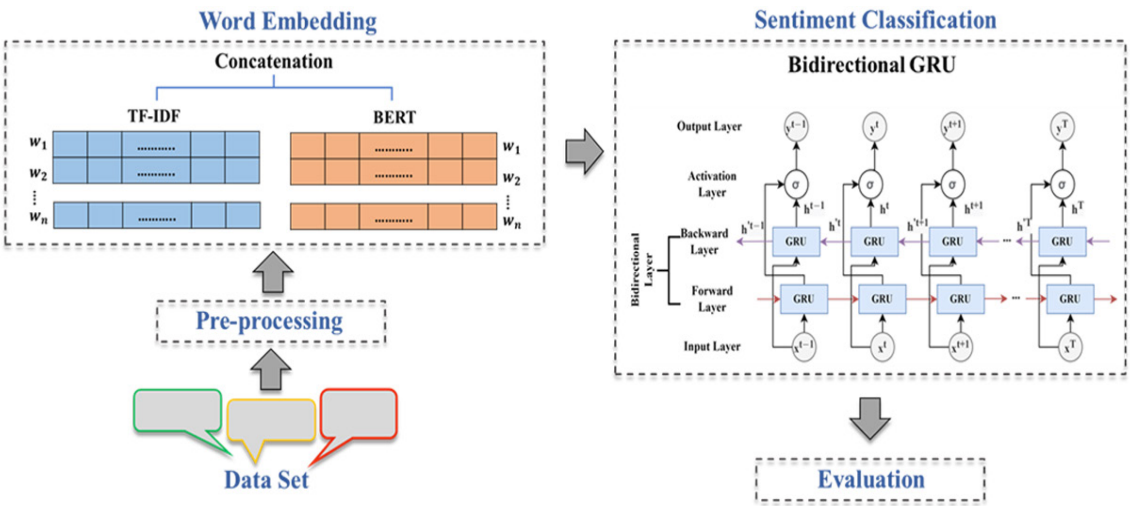

Figure 1: Proposed framework combining Frequency Inverse Document Frequency (TF-IDF) and Bidirectional Encoder Representations from Transformers (BERT) embeddings with a Bidirectional Gated Recurrent Unit (Bi-GRU) classifier for sentiment analysis of Persian healthcare reviews, followed by performance evaluation using standard metrics.

Table 1: Comment lengths in the dataset range from one or two meaningful tokens to a maximum of 91, with representative examples provided in this table.

Comment Length	Sample Comment (Persian)	Sample Comment (English Translation)
One word	معمولی	• ordinary
91 words	من توصیه می‌کنم قبل از شروع درمان، چند بیمارستان را بررسی کنید و تنها پس از آن باید تصمیم گرفت. بر اساس تجربه خودم، نمی‌توانم این بیمارستان را تأیید کنم زیرا هیچ چیز برای بیمار و همراهش مهم‌تر از توجه سریع به نیازهایشان نیست.	• I recommend exploring several hospitals prior to commencing treatment and making a decision only afterward. Based on my own experience, I cannot endorse this hospital because nothing is more significant to a patient and their companion than receiving prompt attention to their needs.

text into sentences and smaller meaningful units, referred to as tokens, based on spacing and punctuation [18].

The stop-word removal step refers to the process of discarding words that are considered insignificant for sentiment analysis and primarily contribute to increased dataset dimensionality due to their high frequency in the text [19]. Stemming reduces words to their base form to minimize lexical redundancy and simplify the dataset [18].

Word Embedding

The meaning of a word can vary depending on the context. For example, the term "shivering" can convey different implications in medical texts. In the sentence "There is no need to worry, and shivering is normal," it refers to a harmless physiological response. In contrast, in "Shivering during fever is caused by the release of cytokines and prostaglandins and should be controlled with medication," it indicates a pathological symptom that requires medical attention. Therefore, accurate sentiment analysis in healthcare comments requires not only understanding medical terminology but also capturing the contextual meaning of words within the text [9, 20-22]. However, traditional embedding and pre-trained models often fail to capture such contextual nuances [9]. Despite fine-tuning on domain-specific corpora, they encounter semantic and syntactic limitations due to insufficient incorporation of external knowledge. To address this, recent studies emphasize integrating context-aware embeddings with structural models that account for word position and co-occurrence [10]. Accordingly, this study employs a hybrid embedding framework that combines BERT and TF IDF. BERT, through its bidirectional transformer architecture and attention mechanism, encodes deep semantic associations and contextual dependencies [23, 24], while TF IDF provides complementary syntactic and statistical term weighting, emphasizing structurally significant words within the dataset [25].

Sentiment Classification

A crucial factor influencing the accuracy of sentence-level sentiment categorization and analysis involves accounting for the sequential order and temporal dependencies among words. Moreover, considering the overall semantic and syntactic context can yield more reliable results in sentiment analysis [9]. Since this study focuses on sentence-level sentiment classification (into three categories: positive, negative, and neutral), the Bi-GRU model is adopted within a deep learning framework. The Bi-GRU architecture, integrating both forward and backward gated recurrent units, is capable of capturing contextual dependencies and temporal relationships between words while maintaining the overall semantic coherence of the text [26, 27]. The principal advantage of the Bi-GRU algorithm lies in its embedded gating mechanism that mitigates information loss across long sequential dependencies [28]. This feature plays a crucial role in preserving essential information, thereby enhancing analytical precision [28]. Furthermore, the Bi-GRU offers a simpler architecture, fewer parameters, and shorter training time than LSTM, Bi-LSTM, and GRU, which share similar structural designs collectively contributing to higher efficiency and improved accuracy in sentiment analysis [23].

Results

The Google Collab environment, Python programming language, and GPU (AMD Radeon HD 7660G) settings were used for implementation.

Pre-processing

Text pre-processing was conducted using the Parsivar library in Python through several steps. First, the Normalizer module standardized comment texts by removing Arabic script and converting colloquial terms to formal equivalents. Next, "Tokenize-Sentences" and "Tokenize-Words" segmented texts into sentences and tokens. Then, a custom Python-based Remove-stop-words function, created using a Persian stop-word list, eliminated non-informative words. Finally, the Stemming module unified different word forms to their roots, reducing redundancy and ensuring consistent lexical representation.

Word Embedding

In this stage, the TF-IDF word embedding model is utilized to analyze the comments and extract syntactic and structural features, producing 300-dimensional vectors. The BERT embedding model is also implemented separately on the dataset to extract semantic features and create 786-dimensional vectors. These vectors are then hybridized and merged using the concatenate technique to create a single 1068-dimensional representation. At this stage, the dimensionality of the embedding vectors for TF-IDF and BERT is set to their standard values of 300 and 768 dimensions, respectively [29, 30].

Sentiment Classification

As indicated by the researchers, when the dataset is divided via the 80-20 or 70-30 split method, it leads to enhanced outcomes for the classification model whilst simultaneously averting overfitting [11]. For this reason, for this step, 70% of the dataset was utilized for training, and the remaining 30% was used for testing. Following that, the Bi-GRU parameters were tuned utilizing the Grid Search method. Afterwards, the classification model was developed utilizing the optimal values (as enumerated in Table 2) and implemented on the dataset. Upon evaluation, as outlined in Table 3, the proposed approach demonstrates accuracy rates amounting to 83.05% and 82.04% F1-score in the classification of Persian healthcare comments into the following three affective categories: negative, positive, and neutral.

Additionally, the performance of the XGBoost, SVM, and Bi-LSTM algorithms was evaluated using the BERT–TF-IDF hybrid embedding model to classify Persian healthcare comments into positive, neutral, and negative categories (as enumerated in Table 4).

Discussion

Effect of Hybrid Embedding on Healthcare Sentiment Accuracy

In this stage, TF-IDF and BERT embeddings

Table 2: Bidirectional (GRU) and dropout rate are internal parameters of the Bi-GRU algorithm. In addition, the loss function, optimizer, metrics, number of epochs, and batch size are the parameters required to build the classification model, for which the optimal values are reported in this table.

	Parameters	Optimal values of parameters
Necessary parameters to create the Bi-GRU model	Bidirectional (GRU)	Units =128
	Bidirectional (GRU)	Units =64
	Dropout rate	0.6
Necessary parameters of model execution	Loss Function	categorical_crossentropy
	Optimizer	Adam
	Metrics	Accuracy F1-Score
	Epochs	70
	Batch size	35

Bi-GRU: Bidirectional Gated Recurrent Unit were implemented separately to evaluate the impact of the proposed hybrid embedding approach in Persian healthcare sentiment analysis. Following pre-processing, TF-IDF and BERT embeddings were independently generated for all comments and used as input to the Bi-GRU sentiment classification model. The Bi-GRU model achieved 63.46% accuracy (62.21% F1) with TF-IDF and 80% accuracy (79.06% F1) with BERT embeddings) as shown in Table 3). In contrast, the proposed hybrid BERT-TF-IDF approach yielded the highest performance, reaching 83.05% accuracy and 82.04% F1-score, confirming its effectiveness in Persian healthcare sentiment analysis.

Although TF IDF reduces the model's processing time compared to other methods, its performance in Persian healthcare sentiment analysis is limited, due to the domain's rich and specialized vocabulary. Because TF IDF depends solely on word frequency, it fails to capture contextual meaning, identify negation, or correctly weight polysemous words [8, 11, 13]. For example, in the sentence "I do not recommend this pill to anyone," the word "recommend" is typically associated with positive sentiment; however, the negation "not"

Table 3: Evaluation of the Bi-GRU sentiment classification model using Precision, Accuracy, Recall, and F1 score for the hybrid BERT + TF-IDF approach and the individual BERT and TF-IDF representations.

	Method		Evaluation Metrics			
	Word embedding	Classification Model	Accuracy	Precision	Recall	F1-score
Proposed method	BERT_ TF-IDF	Bi-GRU	83.05	82	82.09	82.04
Comparative method	TF-IDF	Bi-GRU	63.46	62	62.44	62.21
	BERT	Bi-GRU	80	79	79.19	79.06

BERT: Bidirectional Encoder Representations from Transformers, TF-IDF: Term Frequency Inverse Document Frequency, Bi-GRU: Bidirectional Gated Recurrent Unit

Table 4: The machine learning algorithms XGBoost, SVM, and Bi-LSTM were implemented using the BERT + TF-IDF hybrid model and evaluated on the dataset to compare their performance with the Bi-GRU algorithm and according to the defined criteria.

	Method		Evaluation Metrics			
	Word embedding	Classification Model	Accuracy	Precision	Recall	F1-score
Suggested method for comparison	BERT_ TF-IDF 768+300=1068	XGBoost	79.33	78.51	79.6	79.05
		SVM	76.14	77	77.12	77.05
		Bi-LSTM	82	81	81.21	82.10
Proposed method	BERT_ TF-IDF 768+300=1068	Bi-GRU	83.05	82	82.09	82.04

BERT: Bidirectional Encoder Representations from Transformers, TF-IDF: Term Frequency Inverse Document Frequency, XGBoost: Extreme Gradient Boosting, SVM: Support Vector Machine, Bi-LSTM: Bidirectional Long-Short Term Memory, Bi-GRU: Bidirectional Gated Recurrent Unit reverses the overall sentiment of the sentence. TF-IDF, which relies on surface-level term frequency, cannot model this contextual interaction between words. In contrast, BERT captures contextual meaning and negation through its bidirectional attention mechanism, resulting in more accurate sentiment interpretation [27].

However, the proposed hybrid BERT_TF IDF approach integrates semantic richness from BERT with syntactic and structural weighting from TF IDF, producing more informative feature representations. This combination significantly enhances classification accuracy in healthcare sentiment analysis, while slightly increasing computational time. Although the proposed method slightly increases computational time, it leads to a significant improvement in classification accuracy. The results indicate that this modest increase in processing cost is a reasonable and acceptable trade-off for achieving more reliable sentiment analysis outcomes.

Impact of Classification Models on Healthcare Sentiment Analysis Accuracy

The performance of XGBoost, SVM, and Bi-LSTM algorithms was evaluated using the BERT–TF-IDF hybrid embedding model to classify Persian healthcare comments into positive, neutral, and negative categories. As presented in Table 4, XGBoost obtained 79.33% accuracy and an F1-score of 79.05, whereas SVM yielded 76.14% accuracy and an F1-score of 77.05 under the same hybrid configuration. The superior performance of XGBoost can be attributed to its parameter optimization flexibility, parallel processing, and efficient learning rate control, which enhance both speed and accuracy compared with SVM.

However, these conventional machine learning models cannot explicitly capture sequential dependencies critical to sentence-level sentiment recognition. Deep-learning approaches, particularly Bi-LSTM and Bi-GRU, overcome this limitation by modeling contextual and temporal relationships within healthcare comments. Both architectures exhibit higher predictive power by preserving semantic dependencies and understanding context flow. Among them, Bi-GRU demonstrates slightly higher accuracy and computational efficiency than Bi-LSTM, attributed to its simpler architecture, fewer trainable parameters, and automatic gating system, which mitigates information loss over long sequences. This streamlined structure reduces training time and computational cost while maintaining the ability to encode long-range contextual cues. Consequently, Bi-GRU achieves the highest classification accuracy among all compared models, confirming its advantage for Persian healthcare sentiment analysis tasks involving long and complex textual expressions.

Conclusion

This study introduced a hybrid sentiment analysis framework for Persian healthcare texts that integrates contextual and statistical representations through the combination of BERT and TF IDF embeddings within an adjoint feature representation, followed by a Bi GRU classification model. The motivation behind this approach stems from the complexity of healthcare language, where domain specific terminology, contextual ambiguity, and subtle semantic variations can significantly influence the interpretation of opinions and medical statements.

The experimental evaluation demonstrates that the proposed framework achieves an accuracy of 86% and an F1 score of 84.99%, outperforming models based solely on TF IDF or BERT embeddings. The results confirm that combining contextual embeddings with statistical term weighting enables a richer representation of healthcare texts. The integration of these complementary representations improves the model's ability to identify both domain specific terminology and out of domain expressions while preserving meaningful semantic associations between them.

In addition, the comparative analysis of classification models shows that the Bi GRU architecture provides superior performance compared with XGBoost, SVM, and Bi LSTM. This finding highlights the importance of modeling sequential dependencies and contextual information in sentiment analysis tasks, particularly in the healthcare domain where the meaning and polarity of statements often depend on the broader textual context.

Nevertheless, several limitations remain. The study relies on a specific healthcare dataset, which may restrict the generalizability of the results to other types of clinical or biomedical texts. Moreover, the hybrid embedding strategy introduces additional computational overhead compared with single representation methods. Future work may explore optimizing the computational efficiency of the framework, extending the approach to larger and more diverse healthcare datasets, and incorporating external medical knowledge sources such as clinical ontologies or knowledge graphs to further enhance contextual understanding.

Acknowledgment

The authors gratefully thank all contributors and institutions involved in this study.

Authors' Contribution

F. Foroutan, SR. Khayami, P. Shamsinejad, and R. Javidan contributed to the conceptualization, study design, and methodological development. Data collection, implementation, and drafting were conducted by F. Foroutan, and all authors collaboratively analyzed, revised, and approved the final manuscript.

Ethical Approval

This is an analytical study. No ethical approval is required.

Informed Consent

No direct human interaction or identifiable data were involved; the use of publicly available comments waived the need for informed consent under applicable guidelines.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

None

References

1. NYTLicensing. 29 Healthcare Marketing Statistics To Know for 2024 [Internet]. 2024. Available from: https://nytlicensing.com/latest/trends/healthcare-marketing-stats/.
2. Burstin H, Curry S, Ranney ML, Arora V, Wachler BB, Chou WS, et al. Identifying Credible Sources of Health Information in Social Media: Phase 2-Considerations for Non-Accredited Nonprofit Organizations, For-Profit Entities, and Individual Sources. *NAM Perspect.* 2023;**2023**:10.31478/202305b. doi: 10.31478/202305b. PubMed PMID: 37916063. PubMed PMCID: PMC10617996.
3. RepuGen. RepuGen Patient Review Survey 2024 [Internet]. 2024. [cited 2024 Dec 7]. Available from: https://www.repugen.com/patient-review-survey-2024.
4. ReferralMD. 24 Stats on How Social Media has Impacted the Healthcare Industry [Internet]. 2024. [cited 2024 Dec 7]. Available from: https://referralmd.com/healthcare-social-media-statistics/.
5. Boroumandzadeh M, Parvinnia E, Boostani R, Sefidbakht S. A decision support system framework based on text mining and decision fusion techniques to classify breast cancer patients. *Control Optim Appl Math.* 2021;**6**(1):11-29. doi: 10.30473/coam.2021.60533.1175.
6. Wang Y, Huang G, Li J, Li H, Zhou Y, Jiang H. Refined global word embeddings based on sentiment concept for sentiment analysis. *IEEE Access*. 2021;**9**:37075-85. doi: 10.1109/ACCESS.2021.3062654.
7. Song S, Johnson AP. Predicting drug review polarity using the combination model of multi-sense word embedding and fuzzy latent Dirichlet allocation (FLDA). *IEEE Ac-*

cess. 2023;**11**:118538-46. doi: 10.1109/ACCESS.2023.3326757.

8. Yadav A, Vishwakarma DK. A weighted text representation framework for sentiment analysis of medical drug reviews. In sixth international conference on multimedia big data (BigMM); New Delhi, India: IEEE; 2020. p. 326-32.
9. Khine AH, Wettayaprasit W, Duangsuwan J. A new word embedding model integrated with medical knowledge for deep learning-based sentiment classification. *Artif Intell Med.* 2024;**148**:102758. doi: 10.1016/j.artmed.2023.102758. PubMed PMID: 38325934.
10. Alotaibi FS, Gupta V. Sentiment analysis system using hybrid word embeddings with convolutional recurrent neural network. *Int Arab J Inf. Technol.* 2022;**19**(3):330-5. doi: 10.34028/iajit/19/3/6.
11. Didi Y, Walha A, Wali A. COVID-19 tweets classification based on a hybrid word embedding method. *Big Data Cogn Comput.* 2022;**6**(2):58. doi: 10.3390/bdcc6020058.
12. Duong HT, Nguyen-Thi TA. A review: preprocessing techniques and data augmentation for sentiment analysis. *Comput Soc Netw.* 2021;**8**(1):1-16. doi: 10.1186/s40649-020-00080-x.
13. Rustam F, Khalid M, Aslam W, Rupapara V, Mehmood A, Choi GS. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLoS One.* 2021;**16**(2):e0245909. doi: 10.1371/journal.pone.0245909. PubMed PMID: 33630869. PubMed PMCID: PMC7906356.
14. Bokaee Nezhad Z, Deihimi MA. Twitter sentiment analysis from Iran about COVID 19 vaccine. *Diabetes Metab Syndr.* 2022;**16**(1):102367. doi: 10.1016/j.dsx.2021.102367. PubMed PMID: 34933273. PubMed PMCID: PMC8667351.
15. Taghizadeh N, Doostmohammadi E, Seifossadat E, Rabiee HR, Tahaei MS. SINA-BERT: a pre-trained language model for analysis of medical texts in Persian [Internet]. arXiv [Preprint]. 2021 [cited 2021 April 15]. Available from: https://arxiv.org/abs/2104.07613.
16. W3Techs. Usage Statistics and Market Share of Persian for Websites. 2024. Available from: https://w3techs.com/technologies/details/cl-fa-.
17. Nazarizadeh A, Banirostam T, Sayyadpour M. Sentiment analysis of Persian language: review of algorithms, approaches and datasets [Internet]. arXiv [Preprint]. 2022 [cited 2022 Nov 9]. Available from: https://arxiv.org/abs/2212.06041.
18. Ahuja R, Chug A, Kohli S, Gupta S, Ahuja P. The impact of features extraction on the sentiment analysis. *Procedia Comput Sci.* 2019;**152**:341-8. doi: 10.1016/j.procs.2019.05.008.
19. Pradana AW, Hayaty M. The effect of stemming and removal of stopwords on the accuracy of sentiment analysis on indonesian-language texts. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control.* 2019;**4**(4):375-80. doi: 10.22219/kinetik.v4i4.912.
20. Sanglerdsinlapachai N, Plangprasopchok A, Ho TB, Nantajeewarawat E. Improving sentiment analysis on clinical narratives by exploiting UMLS semantic types. *Artif Intell Med.* 2021;**113**:102033. doi: 10.1016/j.artmed.2021.102033. PubMed PMID: 33685589.
21. Carrillo-de-Albornoz J, Rodríguez Vidal J, Plaza L. Feature engineering for sentiment analysis in e-health forums. *PLoS One.* 2018;**13**(11):e0207996. doi: 10.1371/journal.pone.0207996. PubMed PMID: 30496232. PubMed PMCID: PMC6264154.
22. Asghar MZ, Ahmad S, Qasim M, Zahra SR, Kundi FM. SentiHealth: creating health-related sentiment lexicon using hybrid approach. *Springerplus.* 2016;**5**(1):1139. doi: 10.1186/s40064-016-2809-x. PubMed PMID: 27504237. PubMed PMCID: PMC4954801.
23. Yin X, Liu C, Fang X. Sentiment analysis based on BiGRU information enhancement. *J Phys Conf Ser.* 2021;**1748**(3):032054. doi: 10.1088/1742-6596/1748/3/032054.
24. GeeksforGeeks. Sentiment Classification Using BERT [Internet]. 2024 [cited 2024 May 24]. Available from: https://www.geeksforgeeks.org/sentiment-classification-using-bert/
25. Chiny M, Chihab M, Bencharef O, Chihab Y. LSTM, VADER and TF-IDF based hybrid sentiment analysis model. *Int J Adv Comput Sci Appl.* 2021;**12**(7):265-75. doi: 10.14569/ijacsa.2021.0120730.
26. Shan Y. Social Network Text Sentiment Analysis Method Based on CNN-BiGRU in Big Data Environment. *Mobile Information Systems.* 2023;**2023**(1):8920094. doi: 10.1155/2023/8920094.

27. Li Z, Zhou L, Yang X, Jia H, Li W, Zhang J. User sentiment analysis of Covid-19 via adversarial training based on the BERT-FGM-BiGRU model. *Systems.* 2023;**11**(3):129. doi: 10.3390/systems11030129.

28. Rana MR, Nawaz A, Ali T, El-Sherbeeny AM, Ali W. A BiLSTM-CF and BiGRU-based deep sentiment analysis model to explore customer reviews for effective recommendations. *Eng Technol Appl Sci Res.* 2023;**13**(5):11739-46. doi: 10.48084/etasr.6278.

29. Onan A. Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurrency and computation: Practice and Experience.* 2021;**33**(23):e5909. doi: 10.1002/cpe.5909.

30. Tinkerd. Understanding BERT [Internet]. 2024. [cited 2024 Aug 16]. Available from: https://tinkerd.net/blog/machine-learning/bert-embeddings/.