Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 11, No. 4, 2023, 587-598.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Investigating Shallow and Deep Learning Techniques for Emotion Classification in Short Persian Texts

Mahdi Rasouli and Vahid Kiani*

Department of Computer Engineering, University of Bojnord, Bojnord, Iran.

Article Info

Article History:
Received 29 September 2023
Revised 22 November 2023
Accepted 16 December 2023

DOI:10.22044/jadm.2023.13675.2485

Keywords:
Natural Language Processing, Emotion Classification, Persian Text, Emotion Detection Benchmark, Deep Learning.

**Corresponding author: v.kiani@ub.ac.ir (V. Kiani).*

Abstract

The identification of emotions in short texts of low-resource languages poses a significant challenge, requiring specialized frameworks and computational intelligence techniques. This paper presents a comprehensive exploration of shallow and deep learning methods for emotion detection in short Persian texts. Shallow learning methods employ feature extraction and dimension reduction to enhance classification accuracy. On the other hand, deep learning methods utilize transfer learning and word embedding, particularly BERT, to achieve high classification accuracy. A Persian dataset called "ShortPersianEmo" is introduced to evaluate the proposed methods, comprising 5472 diverse short Persian texts labeled in five main emotion classes. The evaluation results demonstrate that transfer learning and BERT-based text embedding perform better in accurately classifying short Persian texts than alternative approaches. The dataset of this study ShortPersianEmo will be publicly available online at https://github.com/vkiani/ShortPersianEmo.

1. Introduction

In today's world, people increasingly use online social media platforms to communicate their ideas, feelings, and emotions [1]. As a result, sentiment analysis, and emotion recognition using machine learning techniques from text play a crucial role in the automated analysis of individuals and communities [1-4]. This approach facilitates the assessment of the emotional impact of decisions made by industry leaders and policymakers, the determination of customer satisfaction levels in online retail platforms, the analysis of emotions and sentiments expressed by social media users in response to social and political events, evaluation of the emotional state of individuals with psycho disorders during the healing process, exploration of public sentiments and perspectives on specific political figures or governments, as well as the investigation into the effectiveness or ineffectiveness of targeted reform programs concerning social norms and cultural practices [5-7].

Emotion detection in text is a multi-class classification problem in which the intelligent system must recognize human emotion by examining written text [1,8]. An emotion recognition system starts by receiving textual data as input and determines the emotion in the input text using natural language processing and data analysis algorithms. The diagram in Figure 1 illustrates the procedural framework of a text-based emotion recognition system.

This paper presents an investigation into the performance of shallow and deep learning techniques in the domain of emotion detection within short Persian texts. The study introduces two shallow learning methods and two deep learning methods for this task. To assess the effectiveness of these machine learning approaches, a novel Persian dataset called "ShortPersianEmo" is also created and introduced. Notably, this research constitutes the pioneering effort to specifically address emotion detection in short Persian texts.

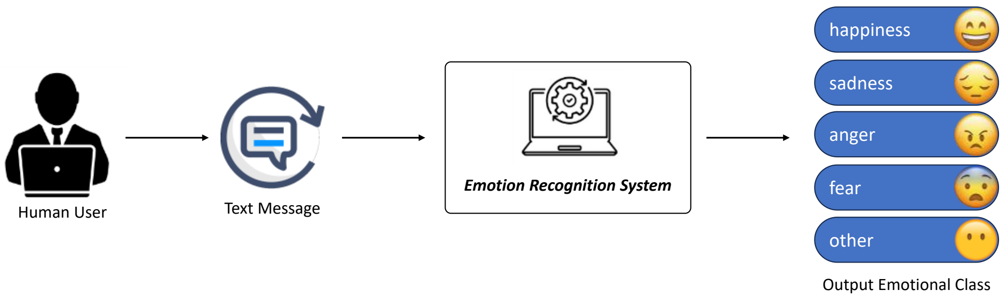

Figure 1. Procedural framework of a text-based emotion recognition system.

This article is organized as follows: Section 2 is dedicated to a review of the previous studies in the field of emotion detection in text. In section 3, shallow and deep learning methods investigated in this research work are described. In section 4, the data collection and labeling are described, and data analysis is presented. The evaluations and results of the experiments are presented in Section 5. Finally, section 6 draws conclusions and suggests future studies.

2. Related Works

Recent studies have witnessed the utilization of deep learning techniques in non-Persian languages to achieve significant accuracy in emotion identification within textual content. Transfer learning and word embedding techniques have demonstrated remarkable efficacy in the domain of emotion recognition, especially when applied to languages with limited resources. In this context, certain scholars have utilized transfer learning to effectively retrain and deploy a pre-existing model that was originally trained for a comparable task, with the objective of discerning emotions expressed within textual data. For instance, the sent2affect approach [3] applies transfer learning and fine-tuning to a model trained for sentiment classification in text, adapting it for emotion recognition. Instead of building a complete emotion recognition model with transfer learning, transfer learning can only be used in the text embedding stage. For example, various versions of BERT, such as Google BERT [9], XLM-RoBERTa [10], and Bangla BERT [11], have been used for text embedding in emotion recognition systems. In addition, in some research works, transfer learning is applied to emotional embedding techniques to augment the comprehension of emotional word semantics within deep learning models. Examples include using Emotional Embedding with FastText embedding [12] and using SSWE emotional embedding with Glove embedding [13]. In conjunction with transfer learning, numerous other methodologies are also employed by researchers to increase the accuracy of emotion identification in textual data. Some researchers combined models from different languages to improve accuracy for languages with limited resources [14]. Data augmentation techniques have been employed to address the challenge of limited training data in languages with scarce resources. Methods such as back translation [15], BART and PREDATOR models [16], C-BERT model [17], and SeqGAN model [18] have been utilized to augment the data for text emotion recognition. A comprehensive overview of textual data augmentation methods can be found in [19,20]. Recently, in [21], the unsupervised zero-shot learning based on a sentence transformer is used to automatically annotate unlabeled data in 34 emotion classes, and then the resulting dataset is used to train a deep learning model.

The field of emotion recognition in Persian text has also seen significant progress in recent years despite limited resources for the language. Early studies, such as the one in [22], relied on the NRC dictionary and shallow learning to identify emotions in Persian text. However, the use of statistical features derived from the dictionary did not yield accurate results. Thus, researchers shifted their focus toward deep learning methods. In [23], a combination of deep learning features and syntactic statistical features was employed to improve emotion identification in Persian texts. This involved the use of Word2Vec word embedding and a GRU recurrent neural network in the machine learning model. Subsequent studies replaced context-free embedding models with BERT-based embedding models, with ParsBERT being one of the most successful models used. In [24], ParsBERT, XLM-R, and XLM-R Large embeddings were applied to the ArmanEmo dataset, with XLM-R Large achieving the highest accuracy. In [25], the same embedding models were combined with data augmentation techniques and applied to the ArmanEmo and EmoPars datasets, with the XLM-R embedding achieving the highest accuracy. Finally, in [26], a deep learning method based on XLM-R embedding was combined with a traditional machine learning model CatBoostDT for emotion recognition on the JAMFA dataset.

The review of existing literature in the field of emotion classification in Persian texts revealed that previous studies have been conducted without taking into account the length of the input text. This observation is made despite the recognized influence that the length of the text can impact the difficulty of its classification for machine learning techniques [27,28]. Classifying short texts is generally more challenging compared to long texts due to several reasons. Firstly, short texts often lower contextual information, making it harder for algorithms to understand the meaning and context of the text. Secondly, short texts have a limited word count, which can result in less information being available for the algorithm to make accurate predictions. Thirdly, when the text contains sarcasm and irony, short texts often contain higher ambiguity. In other words, the inherent brevity of short texts limits the amount of meaningful information available for classification. Advancements in natural language processing have sought to address this issue by incorporating techniques such as word embeddings and transfer learning, which improve the accuracy of classification. Nonetheless, the impact of text length on classification difficulty in the context of emotion classification in Persian texts has been largely overlooked in previous research. Therefore, further investigation is required to understand and address the unique challenges associated with emotion detection in short Persian texts.

3. Proposed Methods

The classification of short texts is generally considered to be more challenging for machine learning techniques compared to the classification of long texts. Our study focuses on investigating machine learning techniques for emotion classification in short Persian texts. These techniques are categorized into two groups: shallow learning and deep learning. In the shallow learning category, we employ Support Vector Machine (SVM) and Random Forest (RF) classification methods, along with feature selection. In the deep learning category, we utilize transfer learning in the embedding phase, employing two methods: FastText [29] and ParsBERT [30] for embedding Persian texts. In this section details of these machine learning techniques are described.

3.1. Shallow Learning

Shallow learning methods commonly employed for emotion classification in text often encompass four main steps: preprocessing, feature extraction, feature selection, and emotion classification. Shallow learning techniques typically rely on statistical analysis for feature extraction, considering the word frequency in the input text and their scarcity in the whole dataset. However, these statistical features fail to capture the sequential order and precedence of words within the text. In this approach, the widely used TF-IDF method is often employed for feature extraction, resulting in a high-dimensional feature space proportional to the number of unique words in the dataset. This poses a challenge for traditional shallow classification models. To address this, a feature selection step is also implemented to reduce the number of input features. Finally, a shallow learning classification technique such as support vector machines, decision trees, random forests, or naïve Bayes is employed for identifying the class of the input text.

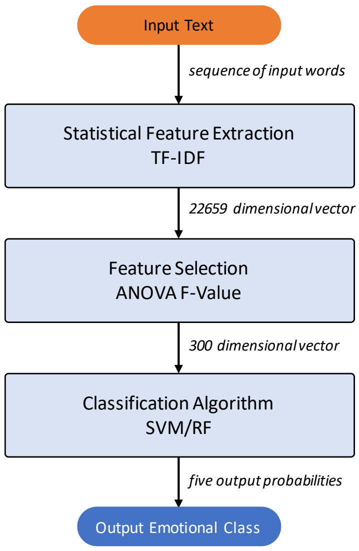

Figure 2. Stages of shallow learning methods investigated for emotion classification in this study.

In this study, we investigate two shallow learning methods that use support vector machines (SVM) and random forest (RF) classification for emotion classification of short Persian texts. The general structure of our proposed shallow learning methods is illustrated in Figure 2. In the feature extraction step, TF-IDF features are extracted from the input texts. Then, in the feature selection step, due to the large number of words and the high dimensions of the created vectors, 300 features were selected from the extracted features using the ANOVA F-Value filter. In the final classification step, these 300 features were fed into traditional SVM and RF classifiers. The characteristics of shallow learning methods evaluated in this study are listed in Table 1. In the Shallow SVM model, we employed the SVM classification method, while in the Shallow RF model, we utilized the Random Forest classification.

Table 1. Shallow learning methods for emotion classification on the ShortPersianEmo dataset.

Model Name	Feature Extraction	Feature Selection	Decision-Making Step
Shallow SVM	TF-IDF	ANOVA-F Value	SVM Classifier
Shallow RF	TF-IDF	ANOVA-F Value	Random Forest Classifier

3.2. Deep Learning

In recent years, there has been a growing interest in utilizing deep learning models to effectively identify emotions in text. Compared to traditional machine learning models, deep learning models have demonstrated superior capabilities in capturing the underlying meaning and intricate patterns of textual data [1,2]. A typical deep learning system for emotion classification in text encompasses four main stages: preprocessing, text embedding, feature learning through deep layers, and decision-making.

Text embedding plays a crucial role in deep learning approaches by transforming the input text into a semantic numerical vector or sequence of vectors. To generate distinct semantic vectors for various words, it is crucial for a text embedding model to undergo training on an extensive corpus of texts. Given the scarcity of textual data in various applications, the integration of transfer learning methodology during the text embedding phase has proven to be beneficial in increasing the accuracy of deep learning models utilized for text classification endeavors.

There are several reasons why transfer learning and text embedding methods are deemed appropriate and effective for short texts. Firstly, emotion classification often requires a large amount of labeled data for training. However, in low-resource languages, obtaining sufficient labeled data can be challenging. Transfer learning and text embedding methods can mitigate this issue by leveraging pre-trained models and word embeddings, reducing the data requirements for training. Secondly, emotion detection in short texts often requires capturing contextual information to accurately classify emotions. Contextual embeddings, such as BERT, capture the meaning of words in the context of the sentence, allowing models to understand the nuances and subtleties of emotions expressed in short texts. Thirdly, deep learning models can be computationally expensive to train from scratch, especially for low-resource languages where limited computational resources are available. Fourthly, pre-trained models and word embeddings are often trained on large and diverse datasets, making them more robust to variations in language, dialects, and writing styles. Fifthly, pre-trained models have undergone training using a vast corpus comprising both lengthy and concise texts. This knowledge transfer enables the generated embedding vectors to offer superior semantic representations for words.

Table 2. Details of the deep learning methods for emotion classification on the ShortPersianEmo dataset.

Model Name	Text Embedding	Embedding Vector Size	Hidden Layers	Decision Layer
Deep FastText	FastText	300	Global Max Pooling	Softmax
Deep ParsBERT	ParsBERT	768	Average Pooling	Softmax

In this study, we investigate two deep models, one based on context-free word embedding, and the other one based on context-based word embedding. These embedding techniques convert text words into numerical vectors. The specifications of these baseline methods based on deep learning are given in Table 2.

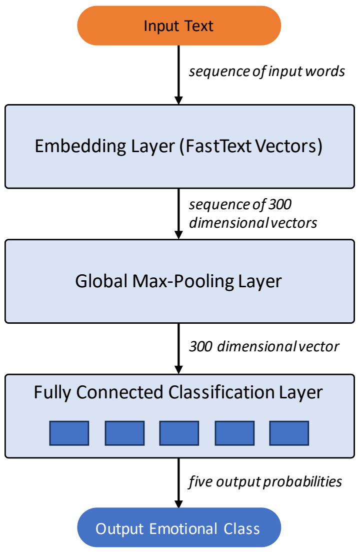

Figure 3. The architecture of the Deep FastText model based on context-free text embedding.

The first deep learning method investigated in this study is based on context-free word embedding and is called Deep FastText. The architecture of this model is depicted in Figure 3. The Deep FastText model is composed of three layers. The initial layer is responsible for the embedding of input text. Following this, a Global Max-pooling layer is employed to transform the matrix of consecutive words into a vector representation. Lastly, a fully connected layer with neurons corresponding to the number of output classes is utilized, incorporating the Soft-max activation function, to make classification decisions. To construct the Deep FastText model, the embedding vectors of length 300 are extracted for all words in the dataset by FastText [29] word embedding which supports Persian[1]. The embedding vectors of FastText are used as the initial weights of the embedding layer in the Deep FastText model. Deep FastText uses transfer learning and fine-tuning techniques on the weights of the embedding layer. At the same time, the weights of other layers are trained during the training process.

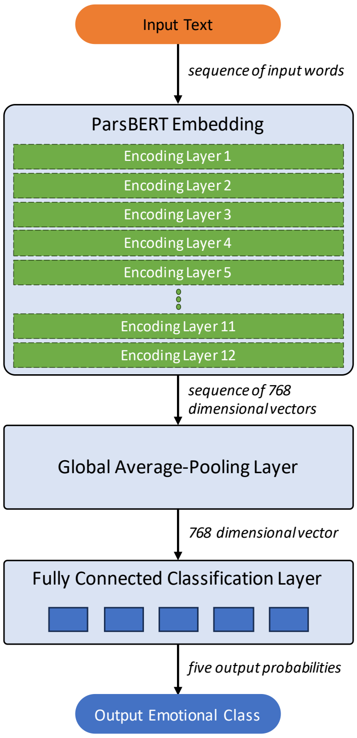

Figure 4. The architecture of the Deep ParsBERT model based on context-sensitive word embedding.

Our second deep model is called Deep ParsBERT which uses ParsBERT pre-trained embedding as a context-sensitive embedding method[2]. ParsBERT is a deep embedding model specifically designed for the Persian language [30]. It is constructed by a combination of bidirectional transformers and a masked language modeling objective to learn contextualized word representations. The ParsBERT model is a deep autoencoder network architecture comprising 12 hierarchical encoding layers. Number of embedding dimensions of ParsBERT for each word is 768. By pretraining on a large corpus of Persian text, ParsBERT can capture the complex syntactic and semantic structures of the language. ParsBERT model can be fine-tuned for downstream tasks such as emotion classification and sentiment analysis. ParsBERT has greatly contributed to advancing research and applications in Persian language processing.

As shown in Figure 4, the proposed Deep ParsBERT model consists of two layers. In the first layer, the input text is transformed into a 768-dimensional vector by ParsBERT embedding. In the second layer, this vector is sent to a fully connected decision layer with neurons equal to the number of output classes and Softmax activation function to perform the classification task. This model uses the advantages of transfer learning and fine-tuning. In the process of training, the weights of the ParsBERT layer will be updated based on the training data, to take advantage of fine-tuning in transfer learning.

The selection of the number of embedding dimensions is a critical hyperparameter in machine learning models utilized for emotion detection in text. This hyperparameter significantly impacts the accuracy of the classification system. A higher value for embedding dimensions introduces complexity to the model, allowing it to represent a greater number of semantic concepts. However, this also increases the risk of overfitting when training data is limited. Conversely, a lower value for embedding dimensions diminishes the model's ability to effectively represent information but reduces the requirement for extensive training data. Therefore, the number of embedding dimensions should be set to a suitable and moderate value. Deep FastText employs pretrained FastText embedding with 300 embedding dimensions, while Deep ParsBERT utilizes pretrained ParsBERT embedding with 768 embedding dimensions. In the case of shallow learning, we employed an empirical setting of 300 features. Thus, Deep ParsBERT uses a higher number of embedding dimensions, can represent a greater number of semantic concepts, and is pretrained on a large amount of text data.

4. Benchmark Dataset

This research presents ShortPersianEmo, a benchmark dataset of short Persian texts labeled with five emotional classes. The selection of short texts for emotion detection was based on some scientific reasons: As each short text typically contains only one emotion, labeling short texts is

[1] https://fasttext.cc/docs/en/crawl-vectors.html

[2] https://huggingface.co/HooshvareLab/bert-base-pars bert-uncased

easier for annotators. This facilitates more accurate evaluation by machine learning methods. By reducing the number of emotions in each text, automatic methods can recognize emotions with higher confidence, leading to more valid results. Finally, the brevity of short texts poses challenges for both readers and machine learning methods, prompting researchers to develop improved methods for emotion classification in this context. Thus, this benchmark dataset provides a more challenging task for researchers and offers an evaluation with higher confidence. This section provides a comprehensive overview of the ShortPersianEmo dataset, offering an in-depth analysis of its key characteristics and features.

4.1. Data Description

The dataset utilized in this study, ShortPersianEmo, is a single-label dataset that contains 5472 short Persian texts collected from Twitter and DigiKala. Our dataset is annotated according to Rachael Jack's emotional model [31] in five emotional classes happiness, sadness, anger, fear, and other.

Table 3. Datasets for emotion detection in Persian texts.

	Emo Pars	**Persian Tweets**	**Arman Emo**	**JAMFA**	**Short Persian Emo***
Instances	29997	113829	7308	2241	5472
Emotion Model	Ekman	Ekman	Ekman	Rachael Jack	Rachael Jack
Emotion Classes	6	6	7	4	5
Labeling	Multi label	Single label	Single label	Single label	Single label
Average Length	113	174	120	NA	56
Min Length	1	6	4	NA	31
Max Length	282	328	423	NA	99

Unlike publicly accessible datasets that do not impose any restrictions on text length, ShortPersianEmo specifically focuses on short texts. The average text length in the ShortPersianEmo dataset is 56 words, while for the EmoPars[3] [32], PersianTweets[4], and ArmanEmo[5] [24], the average text length is 113, 174, and 120 words, respectively. Table 3 presents a comparison between the introduced ShortPersianEmo dataset and other datasets from the literature for emotion detection in Persian text.

4.2. Data Collection

To collect data instances, Twitter and the Digikala website are used as text sources. Data collection from Twitter is done using the Snscrape library in Python programming language. Twitter data gathering is done selectively, and Twitter posts are searched around specific topics. These topics include war, gas, Ukrainian plane, vaccination, Taliban, love, and some other things. For each topic, a specific query is used, which includes some keywords related to the topic. The tweets are collected between 2020 and 2023. At the moment of collection, the length of the texts was not limited, and later texts between 30 and 100 words were selected as short texts. The search was not limited to a specific geographical location, but only Persian tweets were considered in the search. In addition to Persian tweets, our dataset also contains Persian comments from the Digikala website. These comments are selected from a Persian dataset that was previously collected to detect the polarity of opinions in Persian and can be downloaded from the Iranian CamelCase website[6].

4.3. Data Annotation

Each sample in the proposed dataset is labeled with one single label. Emotional labeling of samples is done by five human users. To perform the annotation operation, a web-based system is created using PHP programming language and MySQL database. The collected data are added to the database and re-extracted after annotation.

In our annotation system, a user ID is defined for each annotator. Using the mentioned user ID and provided user interface, each user labeled the instances. After reading each instance, the user specified the emotion through the user interface buttons. Users were encouraged to avoid annotating multi-label or ambiguous instances by clicking on the ignore button. An image of the utilized online annotation system is shown in Figure 5.

4.4. Data Analysis

The ShortPersianEmo dataset contains 5472 instances in five emotion classes happiness, sadness, fear, anger, and other. The number of instances of each class in the ShortPersianEmo dataset is shown in Table 4.

[3] https://github.com/nazaninsbr/persian-emotion-detection
[4] https://www.kaggle.com/datasets/behdadkarimi/persian-tweets-emotional-dataset
[5] https://github.com/arman-rayan-sharif/arman-text-emotion
[6] https://camelcase.ir/

Figure 5. The user interface designed for annotating the dataset.

Table 4. Distribution of emotion classes in the ShortPersianEmo dataset.

Emotional Class	Number of Samples
Happiness	1625
sadness	939
fear	380
anger	1125
other	1403

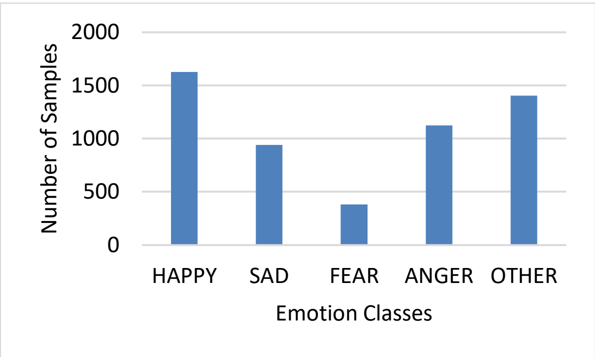

Figure 6. Distribution plot for five emotion classes in ShortPersianEmo dataset.

As it is clear in Table 4, the distribution of data in different classes is imbalanced. The lowest number of samples is for fear class with 380 samples and the highest number of samples is for happiness class with 1625 samples. This imbalance in different classes increases the difficulty of classification for some classification methods. The distribution of instances of each class is also plotted in Figure 6.

Table 5 presents the text and emotional labels of a few instances in the ShortPersianEmo dataset. Emotional labeling of samples is a very difficult task. Because some texts contain several emotions at the same time, the emotional perception of some texts is vague and not the same from the viewpoint of different people. Accordingly, in the labeling of this data set, only text instances were considered that contained only one emotion.

Figure 7 shows the distribution of samples based on the text length between 30 and 100. As it is clear in this figure, the frequency of samples follows a two-phase uniform distribution. Most of the sentences have a short length of between 30 and 50 words, comprising 2783 instances of the entire dataset. The distribution of the frequency in this phase was almost uniform. From a sentence length of 50 words onwards, longer sentences appeared, which suddenly decreased in frequency at each length value. If we consider sentences with a length of 50 words or more as the second phase, the frequency distribution in this phase is also almost uniform. The second phase includes 2689 instances of the entire dataset. The average length of sentences in the entire data set is equal to 56 words.

In Figure 8, word cloud images related to each emotion class in ShortPersianEmo are shown. These images show frequent patterns and words in every emotion. Examining word cloud images for different emotional classes of the proposed data set shows that only paying attention to the appearance of specific keywords is not enough to identify the emotion of each instance. For example, the word "I" was frequent in almost all classes of the data set. The hashtag "Mahsa Amini" has been frequent in all four classes of sadness, fear, anger, and other. Also, the emoji "((" has a high frequency in both classes of fear and other. Hence, discerning the class of every instance within the ShortPersianEmo dataset cannot be solely accomplished through the examination of textual keywords. Accurate classification of this data set necessitates the comprehension of textual semantics through machine learning models.

5. Evaluation and Results

In this section, we will discuss the results of different experiments performed to evaluate the performance of shallow learning and deep learning models on the ShortPersianEmo dataset. The classification performance on short Persian texts of ShortPersianEmo is also compared with the classification performance on longer Persian texts of the ArmanEmo dataset. Finally, in addition to classification performance, the training time for each method will be reported.

Table 5. Some sample instances and their labeling in the ShortPersianEmo dataset.

Text	Class
اگه تو زندگیتون یه رفیق مثل پارسا دارید بله شما خیلی خوشبختید :) English translation: If you have a friend like Parsa in your life, yes, you are very lucky :)	HAPPY
یکی از رفقا بعد از نود و اندی از کسی خوشش اومده و من خوشحال ترینم English translation: One of my friends liked someone after 90 and I am the happiest	HAPPY
وسیله ای بی کیفیت و به درد نخور English translation: A low-quality and useless device	ANGRY
شماهم اینستا میرید عصبی میشید یا فقط من اینطوریم؟؟ English translation: Do you also get nervous when you go to Instagram or is it just me??	ANGRY
این سرفه از هواست یعنی؟ یا از کرونا؟ ترجیح میدم دی اکسید گوگرد تنفس کنم اما کرونا نباشه؟ English translation: Is this cough from the air? Or from Corona? I prefer to breathe sulfur dioxide but not being corona?	FEAR
وای به خدا من میترسم ازاین حیوون English translation: Oh God, I am afraid of this animal	FEAR
راحته ولی کیفیت و زیبایی نداره English translation: It's convenient, but it doesn't have quality or beauty	SAD
از یه جایی به بعد دیگه هیچی مهم نیست English translation: From one point on, nothing matters anymore	SAD
تعطیلی پمپ گازها در پی کمبود گاز در کاشان English translation: Shutdown of gas pumps due to lack of gas in Kashan city	OTHER

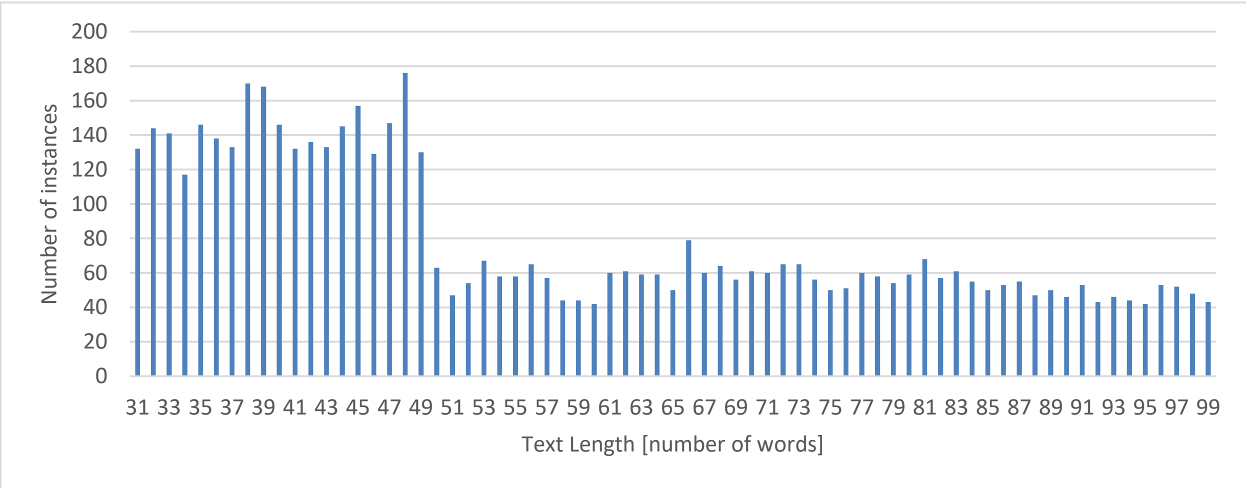

Figure 7. The distribution plot for different text lengths in the ShortPersianEmo dataset.

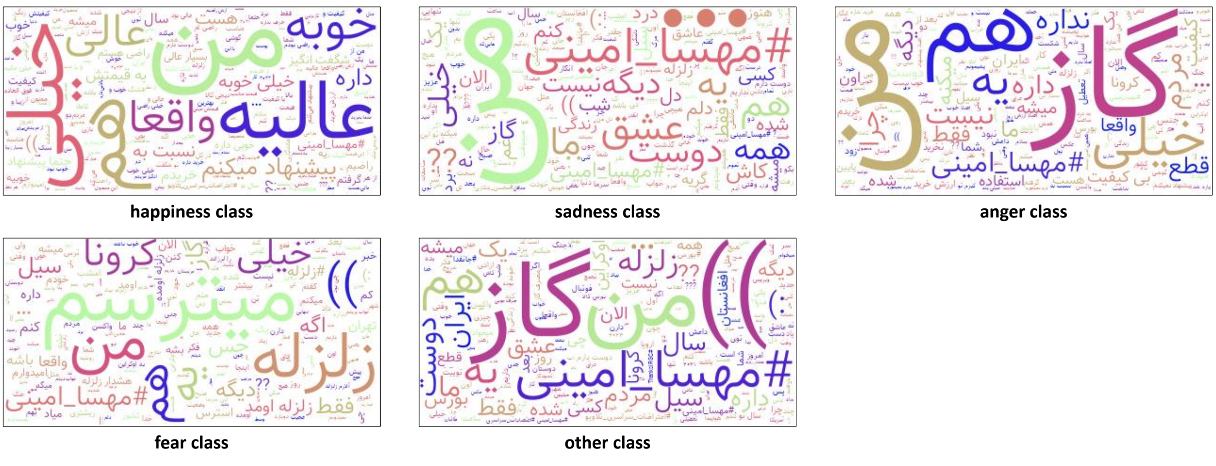

Figure 8. Word clouds of different classes in the ShortPersianEmo dataset.

5.1. Details of simulation

All proposed methods in this study were implemented using Scikit Learn and Keras libraries in Python programming language. The experiments were performed in a Google Colab cloud environment, with 12.7 GB RAM, T4 GPU, and 15.0 GB GPU RAM.

To train the Deep FastText model, the Adam optimizer was used and the batch size was set to 32. To avoid underfitting, overfitting, and long training time, during training of Deep FastText, we also used early stopping with maximum epochs of 100 and patience of 10.

To train the Deep ParsBERT model, the maximum sentence length of the ParsBERT input layer is set to 100 words for the ShortPersianEmo dataset and 128 words for the ArmanEmo dataset. Adam optimizer and batch size equal to 32 are used to train the model. Early stopping is also enabled with maximum epochs of 20 and patience of 2.

Table 6. Distribution of instances for train and test splits.

Model Name	ArmanEmo	ShortPersianEmo
Train	6150	4924
Test	1151	548
All	7301	5472

The assessment of various proposed methodologies is conducted through the implementation of the train-test split technique. Before training the models, the ShortPersianEmo dataset is partitioned into 90% for training purposes and 10% for testing purposes. Similarly, for the ArmanEmo dataset, 85% of the instances are considered for training, whereas 15% of the instances are reserved for testing. The specific distribution of instances for each training and testing split is provided in Table 6 for both datasets.

5.2. Results and Discussion

To evaluate the performance of the proposed methods for emotion detection in short Persian texts, each method was evaluated on the ShortPersianEmo dataset. Additionally, to facilitate a comparative analysis between the emotion classification of short Persian texts and longer Persian texts, the aforementioned methods were also evaluated on the ArmanEmo dataset. The performance of each model on the ShortPersianEmo dataset was then compared to its performance on the ArmanEmo dataset.

In our experiments, the proposed machine learning models were trained and evaluated 10 times on the ShortPersianEmo dataset. The same experiment was also repeated on the ArmanEmo dataset [24]. The average results are summarized in Table 7. In this table, in terms of both Accuracy and Macro-F1 criteria, deep learning-based methods have achieved higher accuracy than shallow learning methods. The transfer learning approach with ParsBERT embedding has achieved the highest accuracy on both datasets in terms of both Accuracy and Macro-F1 criteria. In the Deep ParsBERT method, paying attention to the words before and after each word and considering the context in modeling the meaning of words has made the model have the highest accuracy.

Table 7. Results of shallow and deep learning methods on emotion detection of Persian texts.

	Macro-F1		Accuracy	
Model Name	**Arman Emo**	**Short Persian Emo**	**Arman Emo**	**Short Persian Emo**
Shallow RF	37 %	55 %	37 %	59 %
Shallow SVM	41 %	58 %	51 %	63 %
Deep FastText	42 %	62 %	45 %	65 %
Deep ParsBERT	**65 %**	**71 %**	**67 %**	**73 %**

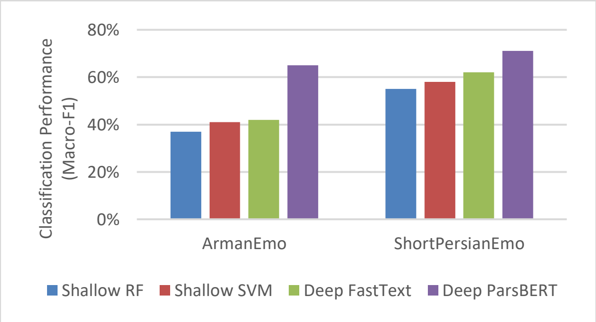

Figure 9. Comparison of the classification accuracy of deep learning models with shallow learning models on emotion detection of Persian texts in terms of Macro-F1 criterion.

Figure 9 shows the accuracy of each of the shallow and deep learning models on the ShortPersianEmo and ArmanEmo datasets in terms of the Macro-F1 criterion. In Figure 9, for the proposed ShortPersianEmo dataset, which contains short Persian texts, Shallow SVM and Shallow RF methods have achieved lower accuracy on average than deep learning models. Nevertheless, the performance of Shallow SVM has been better than Shallow RF. In contrast, deep learning models Deep FastText and Deep ParsBERT have achieved better results. Semantic embedding and transfer learning in the FastText model have made the Deep FastText perform better than the shallow learning models. Finally, the Deep ParsBERT method that uses context-based embedding performed best and provided the highest accuracy on the ShortPersianEmo dataset. The proposed Deep ParsBERT model provides a better performance than the Deep FastText by 9%, compared to the

Shallow SVM by 13%, and compared to the Shallow RF by 16% in terms of the Macro-F1 criterion on the ShortPersianEmo dataset. Examining the results reveals that transfer learning with the help of the ParsBERT model has significantly improved the accuracy of word embedding and emotion classification. This is because words surrounding each word are not taken into account in context-free word embedding, and in shallow learning models that use TF-IDF features.

Figure 9 also presents an opportunity for conducting a comparative analysis by evaluating emotion classification outcomes on two distinct datasets: ShortPersianEmo and ArmanEmo. The results of baseline methods assessed on ShortPersianEmo and ArmanEmo demonstrate that their performance was better in classifying short text data than classifying longer text data, potentially due to the reduced ambiguity in labeling short text data. The Deep ParsBERT method achieved the highest classification accuracy on both datasets thanks to context-sensitive embedding and pretraining on a large text corpus.

Table 8. The training time required for each method on the Persian text datasets (in seconds).

Model Name	Arman Emo	Short Persian Emo
Shallow RF	5.43	3.09
Shallow SVM	**4.76**	**2.07**
Deep FastText	719.00	594.96
Deep ParsBERT	1121.75	908.65

Table 8 presents the training time required for different models on Persian text datasets, namely ArmanEmo and ShortPersianEmo. The training times are measured in seconds. It can be observed that the deep learning models, Deep FastText and Deep ParsBERT, require significantly more training time compared to the shallow methods. Specifically, Deep FastText takes approximately 719.00 seconds for ArmanEmo and 594.96 seconds for ShortPersianEmo, while Deep ParsBERT takes around 1121.75 seconds for ArmanEmo and 908.65 seconds for ShortPersianEmo. On the other hand, the shallow methods, Shallow RF and Shallow SVM, exhibit considerably shorter training times, with Shallow RF requiring approximately 5.43 seconds for ArmanEmo and 3.09 seconds for ShortPersianEmo, and Shallow SVM requiring roughly 4.76 seconds for ArmanEmo and 2.07 seconds for ShortPersianEmo. In the context of Deep FastText, a substantial portion of the training duration is allocated to the process of word retrieval from the FastText dictionary and subsequent replication of the corresponding embedding vector into the initial embedding matrix of the proposed model. Conversely, in the case of the Deep ParsBERT, the predominant share of the training time can be attributed to the computational operations necessary for fine-tuning the weights of the large and complex embedding layer constructed from ParsBERT. While the training process of deep learning-based approaches necessitates a substantial amount of time, the prediction phase is fast and prediction time is tolerable. Conversely, deep learning methods have exhibited highly advantageous outcomes in accuracy.

6. Conclusion

This research article presented an investigation into emotion classification in short Persian texts, employing a range of shallow and deep learning approaches. To facilitate this study, a dataset consisting of short Persian texts was precisely collected and annotated with five distinct emotion classes. The evaluation of both shallow and deep learning models on the ShortPersianEmo dataset showcased the superior accuracy of deep learning models, particularly those leveraging semantic embedding and transfer learning techniques, in effectively classifying emotions. Notably, the deep learning model incorporating ParsBERT emerged as the most successful in achieving exceptional performance in Persian emotion recognition on short texts. Furthermore, a comparative analysis of model outcomes on the ShortPersianEmo and the ArmanEmo datasets revealed that machine learning models displayed enhanced accuracy in classifying short texts as opposed to longer texts, primarily due to the decreased ambiguity associated with labeling shorter text instances. Future research endeavors may concentrate on exploring novel approaches to further enhance the comprehension of the semantic nuances inherent in short texts.

References

[1] S. Peng et al., “A survey on deep learning for textual emotion analysis in social networks,” *Digital Communications and Networks*, vol. 8, no. 5, pp. 745-762, Oct. 2022.

[2] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, “Text-based emotion detection: Advances, challenges, and opportunities,” *Engineering Reports*, vol. 2, no. 7, p. e12189, 2020.

[3] B. Kratzwald, S. Ilić, M. Kraus, S. Feuerriegel, and H. Prendinger, “Deep learning for affective computing: Text-based emotion recognition in decision support,”

Decision Support Systems, vol. 115, pp. 24-35, Nov. 2018.

[4] A. Lakizadeh and E. Moradizadeh, "Text Sentiment Classification based on Separate Embedding of Aspect and Context," *Journal of AI and Data Mining*, vol. 10, no. 1, pp. 139-149, Jan. 2022.

[5] V. Balakrishnan and W. Kaur, "String-based Multinomial Naïve Bayes for Emotion Detection among Facebook Diabetes Community," *Procedia Computer Science*, vol. 159, pp. 30-37, Jan. 2019.

[6] S. Ghosh, N. Froelich, and C. Aragon, "'I Love You, My Dear Friend': Analyzing the Role of Emotions in the Building of Friendships in Online Fanfiction Communities," in *Social Computing and Social Media*, A. Coman and S. Vasilache, Eds., in *Lecture Notes in Computer Science*, vol. 14026. Cham: Springer Nature Switzerland, 2023, pp. 466-485.

[7] M. Nasiri and H. Rahmani, "DENOVA: Predicting Five-Factor Model using Deep Learning based on ANOVA," *Journal of AI and Data Mining*, vol. 9, no. 4, pp. 451-463, Nov. 2021.

[8] M. Rasouli and V. Kiani, "A survey on deep learning methods for text-based emotion classification: Advances, challenges, and opportunities," *Soft Computing Journal*, Sep. 2023, doi: 10.22052/scj.2023.248812.1126.

[9] P. Kumar and B. Raman, "A BERT based dual-channel explainable text emotion recognition system," *Neural Networks*, vol. 150, pp. 392-407, Jun. 2022.

[10] S. Q. Suidong Qu, Y. Y. Yanhua Yang, and Q. Q. Qinyu Que, "Emotion Classification for Spanish with XLM-RoBERTa and TextCNN.," in *Iberian Languages Evaluation Forum 2021 & 37th International Conference of the Spanish Society for Natural Language Processing*, 2021, pp. 94-100. [Online]. Available: https://ceur-ws.org/Vol-2943/emoeval_paper10.pdf

[11] A. Das, O. Sharif, M. M. Hoque, and I. H. Sarker, "Emotion Classification in a Resource Constrained Language Using Transformer-based Approach," arXiv:2104.08613, Apr. 2021. [Online]. Available: https://arxiv.org/abs/2104.08613. [Accessed Oct. 20, 2023].

[12] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-Emotion Neural Network for Emotion Recognition From Text," *IEEE Access*, vol. 7, pp. 111866-111878, 2019.

[13] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, "Understanding Emotions in Text Using Deep Learning and Big Data," *Computers in Human Behavior*, vol. 93, pp. 309-317, Apr. 2019.

[14] M. Abdullah, M. Hadzikadicy, and S. Shaikhz, "SEDAT: Sentiment and Emotion Detection in Arabic Text Using CNN-LSTM Deep Learning," in *17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando*, FL, 2018, pp. 835-840.

[15] H. Luo, "Emotion Detection for Spanish with Data Augmentation and Transformer-Based Models," in *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021), Malaya*, Spain, 2021, pp. 35-42.

[16] H. Q. Abonizio, E. C. Paraiso, and S. Barbon Junior, "Toward Text Data Augmentation for Sentiment Analysis," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 5, pp. 657-668, 2021.

[17] X. Wu, S. Lv, L. Zang, J. Han, and S. Hu, "Conditional BERT Contextual Augmentation," in *Computational Science – ICCS 2019*, J. Rodrigues et al., Eds., in *Lecture Notes in Computer Science*, vol. 11539. Cham: Springer International Publishing, 2019, pp. 84-95.

[18] J. Luo, M. Bouazizi, and T. Ohtsuki, "Data Augmentation for Sentiment Analysis Using Sentence Compression-Based SeqGAN With Data Screening," *IEEE Access*, vol. 9, pp. 99922-99931, 2021.

[19] S. Y. Feng, V. Gangal, J. Wei, S. Chandar, S. Vosoughi, T. Mitamura, E. Hovy, "A Survey of Data Augmentation Approaches for NLP," arXiv: 2105.03075, Dec. 2021. [Online]. Available: https://arxiv.org/abs/2105.03075. [Accessed Oct. 20, 2023].

[20] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 8, no. 1, Art. no. 1, Dec. 2021.

[21] S. G. Tesfagergish, J. Kapočiūtė-Dzikienė, and R. Damaševičius, "Zero-Shot Emotion Detection for Semi-Supervised Sentiment Analysis Using Sentence Transformers and Ensemble Learning," *Applied Sciences*, vol. 12, no. 17, Art. no. 17, Jan. 2022.

[22] A. Khosravi, M. Kelarestaghi, and M. Purmohammad, "Emotion Detection in Persian Text: A Machine Learning Model," *Contemporary Psychology, Biannual Journal of the Iranian Psychological Association (BJCP)*, vol. 14, no. 1, pp. 42-48, Aug. 2019.

[23] S. S. Sadeghi, H. Khotanlou, and M. Rasekh Mahand, "Automatic Persian Text Emotion Detection using Cognitive Linguistic and Deep Learning," *Journal of AI and Data Mining*, vol. 9, no. 2, pp. 169-179, Apr. 2021.

[24] H. Mirzaee, J. Peymanfard, H. H. Moshtaghin, and H. Zeinali, "ArmanEmo: A Persian Dataset for Text-based Emotion Detection," arXiv:2207.11808, Jul. 2022. [Online]. Available: https://arxiv.org/abs/2207.11808. [Accessed Oct. 20, 2023].

[25] A. Abaskohi, N. Sabri, and B. Bahrak, "Persian Emotion Detection using ParsBERT and Imbalanced Data Handling Approaches," arXiv:2211.08029, Nov. 2022. [Online]. Available: https://arxiv.org/abs/2211.08029. [Accessed Oct. 20, 2023].

[26] A. Khodaei, A. Bastanfard, H. Saboohi, and H. Aligholizadeh, “Deep Emotion Detection Sentiment Analysis of Persian Literary Texts,” Research Square, 2022. [Online]. Available: https://www.researchsquare.com/article/rs-1796157/v1. [Accessed Oct. 20, 2023].

[27] Y. Liu, P. Li, and X. Hu, “Combining context-relevant features with multi-stage attention network for short text classification,” *Computer Speech & Language*, vol. 71, p. 101268, Jan. 2022.

[28] J. Xu et al., “Incorporating context-relevant concepts into convolutional neural networks for short text classification,” *Neurocomputing*, vol. 386, pp. 42-53, Apr. 2020.

[29] T. Mikolov, E. Grave, P. Bojanowski, C. Puhrsch, and A. Joulin, “Advances in Pre-Training Distributed Word Representations.” arXiv: 1712.09405, Dec. 2017. [Online]. Available: https://arxiv.org/abs/1712.09405. [Accessed Oct. 20, 2023].

[30] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, “ParsBERT: Transformer-based Model for Persian Language Understanding,” *Neural Processing Letters*, vol. 53, no. 6, pp. 3831-3847, Dec. 2021.

[31] R. E. Jack, W. Sun, I. Delis, O. G. B. Garrod, and P. G. Schyns, “Four not six: Revealing culturally common facial expressions of emotion,” *Journal of Experimental Psychology: General*, vol. 145, no. 6, pp. 708-730, Jun. 2016.

[32] N. Sabri, R. Akhavan, and B. Bahrak, “EmoPars: A Collection of 30K Emotion-Annotated Persian Social Media Texts,” in *Proceedings of the Student Research Workshop Associated with RANLP 2021*, 2021, pp. 167-173. [Online]. Available: https://aclanthology.org/2021.ranlp-srw.23/. [Accessed Oct. 20, 2023].

بررسی تکنیک‌های یادگیری عمیق و کم عمق برای طبقه‌بندی هیجان در متون کوتاه فارسی

مهدی رسولی و وحید کیانی*

گروه مهندسی کامپیوتر، دانشگاه بجنورد، بجنورد، ایران.

ارسال ۲۰۲۳/۰۹/۲۹؛ بازنگری ۲۰۲۳/۱۱/۲۲؛ پذیرش ۲۰۲۳/۱۲/۱۶

چکیده:

شنا سایی هیجان در متون کوتاه زبان‌های کم منبع چالش مهمی ا ست که نیازمند چارچوب‌های خاص و تکنیک‌های هوش محا سباتی ا ست. این مقاله کاوشی جامع بر روش های یادگیری عمیق و کم عمق را برای تشخیص هیجان در متون کوتاه فارسی ارائه می‌کند. روش های یادگیری کم عمق در این مطالعه، از استخراج ویژگی و کاهش ابعاد برای افزایش دقت طبقه‌بندی استفاده می‌کنند. از سوی دیگر، روش های یادگیری عمیق در این بررسی از یادگیری انتقالی و جاسازی کلمه، به ویژه جاسازی BERT برای دستیابی به دقت طبقه‌بندی بالا استفاده می‌نمایند. در این مقاله، همچنین یک مجموعه داده فارسی به نام "ShortPersianEmo" برای ارزیابی روش‌های پیشنهادی معرفی شده است. این مجموعه داده شامل ۵۴۷۲ متن کوتاه فارسی متنوع است که در پنج کلاس هیجانی برچسب گذاری شده‌اند. نتایج ارزیابی نشان می‌دهد که یادگیری انتقالی و جاسازی متن مبتنی بر BERT در طبقه‌بندی هیجانی متون کوتاه فارسی نسبت به رویکردهای جایگزین دقت بالاتری دارند. مجموعه داده ShortPersianEmo به صورت آنلاین در https://github.com/vkiani/ShortPersianEmo در دسترس عموم قرار گرفته است.

کلمات کلیدی: پردازش زبان طبیعی، طبقه‌بندی هیجان، متن فارسی، مجموعه‌داده تشخیص هیجان، یادگیری عمیق.