Iranian Journal of Operations Research
Vol. 16, No. 1, 2025, pp. 136-146

Emotion Recognition in Persian Texts Using an Improved Transformer Model

Elham Askari[1]

Emotion recognition in Persian texts using data mining is a significant area within text analysis. Emotions are typically defined as individuals' emotional reactions to situations, events, and information. Emotion recognition in text involves identifying and analyzing emotional content across various types of textual data. This paper presents a model for detecting different emotions in Persian texts using an improved transfer learning model. The proposed model comprises an encoder and a decoder, each equipped with a self-attention mechanism and RNN modules. First, a dataset of sentences annotated with emotional states, anger, happiness, sadness, and fear is constructed through user contributions. These sentences are then converted into image representations and fed into enhanced transfer learning model for emotion recognition. Experimental results demonstrate that the model effectively identifies the emotions of anger, happiness, sadness, and surprise, precision, recall, accuracy, F1-measure and specificity of 90.25%, 91.6%, 91.4%, 90.80% and 90.31% respectively.

***Keywords**: Persian Texts, Transformer Model, Emotion, RNN*

1. Introduction

Emotions are complex phenomena characterized by both physical and psychological changes that influence an individual's thoughts and behaviors. They encompass physiological arousal, specific behavioral responses, and conscious experiences [22]. Emotions have long been considered a key aspect of human behavior, with their study dating back to the early development of scientific psychology.

In recent years, emotion recognition in text has gained increasing attention in artificial intelligence research, particularly in the development of human-computer interaction systems [16]. Understanding the relationship between text and emotion—how textual content influences readers emotionally or conveys the author's emotional state,is a prominent research area with a growing body of literature [20].

The advent of computational tools has enabled the systematic analysis of large volumes of textual data. Many software applications have been developed for this purpose. Emotion recognition in text, particularly textual content shared on social media and online platforms, provides valuable insights into how individuals respond emotionally to various topics or events [3]. For instance, modern search engines can analyze emotional tones in song lyrics and recommend songs based on the listener's mood [25].

According to linguistic theories, emotion recognition in text can be performed using textual cues. The collection of cues and structures associated with emotions is referred to as "emotional language", which includes both explicit and implicit signals. Explicit cues involve emotion-related words or symbols, which can be classified as either expressive (direct) or descriptive (indirect). Beyond

[1] Corresponding Author.
[1] Department of Computer Engineering, FSh.C., Islamic Azad University, Fouman, Iran: askary.elham@iau.ir

[Downloaded from iors.ir on 2026-11-27]

vocabulary, emotions can also be conveyed through metaphors, phonetic markers, and morphological patterns [10].

By leveraging techniques from computer science and computational linguistics, particularly natural language processing (NLP), numerous systems have been developed for automatic text processing. However, differences in language characteristics, such as phonetic features, syntax, semantics, and idiomatic expressions, pose challenges for cross-linguistic emotion recognition. Effective emotion detection models must draw on both emotion theories and NLP methodologies. Research has shown that a combination of lexical, syntactic, and semantic features can enable models to distinguish emotional from non-emotional content [12, 21].

Several annotated corpora and emotion-based lexical resources have been developed for different languages, such as LIWC [11], the LEW List [23], and WordNet-Affect [2]. These dictionaries associate words not with their literal meanings but with the emotions they convey. However, each resource has its own strengths and limitations. For example, some provide emotion labels without distinguishing between different meanings or senses of a word.

Given the vast amount of textual data available online, emotion recognition in text has become increasingly important, yet challenging. Human emotions are often subtle and not always explicitly expressed. Individuals may use language that does not accurately reflect their true feelings, and writing styles vary significantly between users. These factors complicate emotion recognition tasks.

In this study, we aim to achieve high levels of accuracy and recall by integrating cognitive and neural network-based features. Our proposed model takes into account emotion-laden constructs and keywords, leveraging neural network architectures to enhance performance. Emotional content in text can enrich the reader's connection with a narrative or topic, encouraging deeper emotional engagement. A well-crafted emotional message can resonate with readers, making content more impactful and memorable [24].

To automatically classify emotions in text, contextual features, annotated corpora, emotion lexicons, and computational models are employed. Creating comprehensive and high-quality emotion lexicons is essential for accurate emotion classification [19, 1]. Several efforts have been made to develop such resources, including labeled word lists and annotated corpora in various languages. For instance, LIWC, LEW List, and WordNet-Affect associate words with emotion types rather than meanings. Importantly, emotion analysis can be performed at multiple levels, individual words, sentences, or entire documents [11]. Thus, training data and annotation schemes must account for these levels of granularity.

Common approaches for text-based emotion recognition include [10,12,21,25]:

1. Feature-based methods: These involve extracting features such as keywords, specific phrases, word frequency, and the presence of symbols or emojis. Machine learning algorithms like artificial neural networks (ANNs) and support vector machines (SVMs) are then used to classify emotions.
2. Lexicon-based methods: These rely on dictionaries that associate words and phrases with emotional labels. Using such resources, emotions are detected and assigned to the text accordingly.
3. Deep learning-based methods: Advanced models such as recurrent neural networks (RNNs) and transformers are utilized to capture deeper semantic and contextual information, resulting in more accurate emotion recognition [13].

Emotion detection in Persian text introduces unique challenges, such as cultural nuances and language-specific expressions. High-quality training data and attention to these linguistic and cultural details are essential for successful data mining in Persian [9].

[Downloaded from iors.ir on 2026-11-27]

Processing Persian also presents technical challenges, such as the complexity of its script. For example, the incorrect use of space in place of half-space characters can affect text interpretation. Proper use of Persian punctuation and avoiding the use of Arabic characters in Persian writing can reduce these issues. Additional challenges include the variable forms of compound words, which must be accurately identified and processed [19].

Despite numerous advancements in emotion recognition using deep learning and transfer learning models for widely used languages like English and Chinese, there remains a significant lack of research focused on Persian texts. Persian language processing faces unique challenges, including script complexity, the use of half-spaces, and culturally specific expressions, which are not well addressed by existing models trained predominantly on other languages. Additionally, publicly available annotated datasets for emotion recognition in Persian are scarce and limited in size, restricting the development and evaluation of robust models. Most existing methods rely heavily on text-based features without exploring multi-modal or image-based representations of text, which could capture spatial and structural features useful for emotion detection. This study addresses these gaps by proposing a novel model that leverages image-based representations of Persian texts in combination with transformer and RNN architectures, specifically designed to capture the nuanced emotional content in Persian language. In this paper, we utilize a transfer learning model for emotion detection in Persian texts. The primary advantage of using transfer models lies in their reduced reliance on emotion-specific training data. However, transfer learning has limitations. Emotional patterns in the pre-trained source data may differ from those in the target data, and issues such as "negative transfer" can arise, in which irrelevant features are transferred, resulting in reduced accuracy. Nevertheless, transfer learning remains a promising approach by enabling the reuse of knowledge from large datasets. For better performance, fine-tuning with more targeted emotion-specific data is recommended.

This paper is organized as follows: Section 2 presents the literature review. Section 3 introduces the materials and methods, including the proposed model. Section 4 presents the experimental results.

2. Materials and Proposed Method

In this study, we present a model for recognizing various emotions in Persian texts using a transfer learning approach. The proposed system is based on an improved transfer learning architecture, incorporating both encoder and decoder components with an integrated attention mechanism.

To begin, a dataset consisting of 100 Persian-language sentences is compiled. These sentences are categorized and annotated according to four emotional states: anger, happiness, sadness, and surprise. The annotation process is carried out with the assistance of multiple human contributors to ensure reliability and reduce subjective bias. Next, the annotated sentences are converted into image representations to be used as input for the proposed model. The use of image-based representations allows the model to leverage spatial patterns in text layout and enhances compatibility with deep learning architectures that process visual features. Figure 1 illustrates the block diagram of the proposed method.

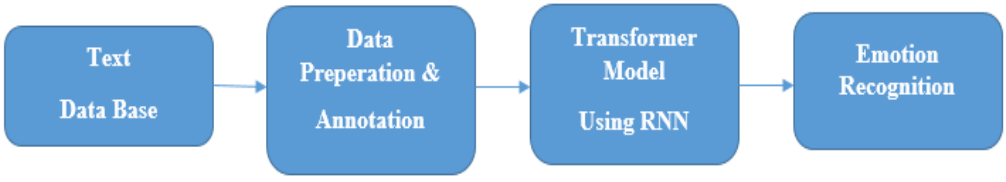

Figure 1. The block diagram of the proposed method

[Downloaded from iors.ir on 2026-11-27]

The following sections provide a detailed explanation of each step.

2.1. Text Database

A visual database containing 100 Persian sentences representing the emotions of anger, happiness, sadness, and surprised are used.

2.2. Data Preparation

The process of converting raw data into a format that is readable and understandable by machines is **known as** data preparation. To perform any activity related to texts, it is necessary to first clean and preprocess the **text**. In this paper, text normalization, converting text to sentences, converting sentences to words, **rooting**, recognizing the role of words, and segmentation will be performed [17,18,27]. These steps are explained in detail below.

a) Normalization

In the normalization process, we want all texts to be standardized and unified. For example, converting English numbers to Persian **(e.g., "1" to "۱")**, removing spaces and extra blank spaces, removing accents from words, and **other transformations as required to ensure text uniformity** [6].

b) Converting text to sentences

During this process, a text is divided into a number of sentences. The division criterion is usually the sentence-ending symbols such as ".!?".

c) Converting sentences to words

In this process, the sentence is converted into a series of words (**tokens**) [5].

c) Stemming

In the stemming process, the root of a word is obtained by applying a series of algorithmic steps. Typically, the algorithm first removes prefixes, then suffixes, proceeding step-by-step to extract the stem.

2.3. Annotation

Text annotation is important because it helps bridge the gap between unstructured text data and structured, machine-readable data. It allows machine learning models to learn patterns from annotated examples and generalize. High-quality annotation is crucial for building accurate and robust models. This is why careful attention to detail, consistency, and domain expertise are essential in text annotation. [5,6,26,28].

3. Proposed Method Based on Transformer Learning

Transfer Learning is a machine learning approach in which a developed model is used to perform new tasks. This technique is a popular approach in deep learning, in which pre-trained models are used as a starting point for problems based on Natural Language Processing [8]. Transfer learning is a type of optimization that accelerates learning or enhances performance on new tasks by leveraging knowledge from related tasks. This method is generally used in situations where there is little data available to model a new phenomenon [7].

In this paper, using the transfer learning technique, the initial and middle layers of the model are reused, and only the final layers are retrained to achieve the desired output in the task that the model is working on. The most important advantages of using the transfer learning technique include greatly

[Downloaded from iors.ir on 2026-11-27]

reduced training time, reduced need for a large amount of data, and improved neural network performance in most cases [4,7,8]. In this paper, the goal is to classify emotions in Persian text, and a transformer model with a self-attention mechanism is used. Combining RNN with a transformer model can enhance the extraction of sequential linguistic features and improve performance. Figure 2 illustrates the transfer learning model used.

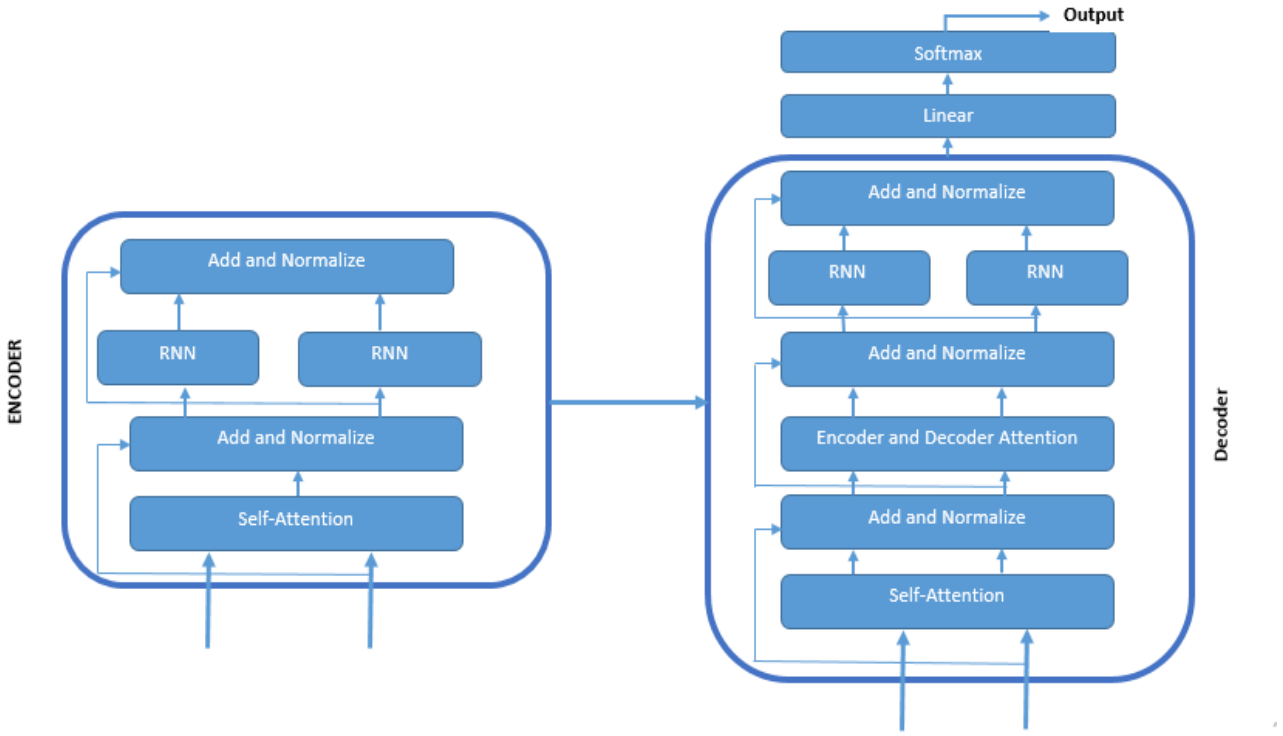

Figure 2. The structure of proposed method

The transformer model used consists of two main parts:

1. **Encoder:** The part that receives the input and extracts its features.
2. **Decoder:** The part that converts the extracted features into the final output.

Each Encoder and decoder consists of several layers, each layer consisting of two main components:

- Self-Attention Mechanism, this mechanism allows the model to give different weights to the input words and examine the relationships between them. Equation (1-8) shows the Self-Attention formula. The vectors Q, K and V are Query, Key and Value, respectively. X is the input. i and j indicate the position in each matrix. W is the weight matrix. μ is the mean and σ is the standard deviation.

$$Q^{(h)} = W_{h,q}{}^{T} X_i, \quad K^{(h)}(X_i) = W_{h,k}{}^{T} X_i, \quad V^{(h)}(X_i) = W_{h,v}{}^{T} X_i,$$
$$W_{h,q}, W_{h,k}, W_{h,v} \in R^{d \times k} \tag{1}$$

$$a_{i,j}{}^{(h)} = soft\max_j \left(\frac{(Q^{(h)}(X_i), K^{(h)}(X_j))}{\sqrt{k}}\right) \tag{2}$$

[Downloaded from iors.ir on 2026-11-27]

$$u_i^{'} = \sum_{h=1}^{H} W_{c,h}^{T} \sum_{j=1}^{n} a_{i,j}^{(h)} V^{(h)}(X_j) \quad W_{c,h} \in R^{K \times d} \tag{3}$$

$$u_i = LayerNorm(X_i + u_i^{'}; \gamma_1, \beta_1) \quad \gamma_1, \beta_1 \in R^{K \times d} \tag{4}$$

$$z_i^{'} = W_2^{T} \operatorname{Re} LU(W^{T}{}_1 U_i), \quad W_1 \in R^{d \times m}, W_2 \in R^{m \times ds} \tag{5}$$

$$Z_i = LayerNorm(u_i + z_i^{'}; \gamma_2, \beta_2) \quad \gamma_2, \beta_2 \in R^{d} \tag{6}$$

$$LayerNorm(z; \gamma, \beta) = \gamma \frac{(z - \mu_z)}{\sigma_z} + \beta, \quad \gamma, \beta \in R^{K} \tag{7}$$

$$\mu_z = \frac{1}{k} \sum_{i=1}^{k} z_i, \quad \sigma_z = \sqrt{\frac{1}{k} \sum_{i=1}^{k} (z_i - \mu_z)^2} \tag{8}$$

The neural network used is an RNN that is independent for each input position. Each of the Decoder and Encoder has an input attention mechanism. This attention mechanism allows the model to pay attention to features. The parameter Θ, consists of the inputs of the weight matrices W. Together with the parameters LayerNorm, γ and β, it is shown on the right. The input $x \epsilon R^{n \times d}$, its similar structure should be interpreted as a set of n objects, each with d features (often, but not always, a length and sequence of vectors). The output $z \epsilon R^{n \times d}$, has a similar structure to the input $x \epsilon R^{n \times d}$. A transformer is a combination of L-shaped transformer blocks. Each with its own parameters $f_{\Theta L}$... $f_{\Theta L}(x) \epsilon R^{n \times d}$. The parameters of the supertransformer are d, k, m, H and L. The hyperparameter settings are d=512, k=64, m=2048, H=8. In the original paper, L=6 is set. n is the number of objects.

After each Attention layer, an RNN with 128 hidden layers will be used. Dropout is performed at a rate of 0.3 to prevent overfitting. The final output of the Decoder is transformed into a probability distribution for each word in the dictionary through a Softmax layer [4,7,8].

4. Experimental Results

The Python programming language was used to implement the proposed method. According to the proposed model, a recurrent neural network was employed within the transfer learning framework. In addition to **dropout**, the k-fold method, with k=10, was used to train and validate the proposed method. The k-fold method is a common technique for evaluating and training artificial intelligence models. This method randomly divides the training data into k parts, selects one part as the validation set at each stage, and uses the remaining parts to train the model. This process is repeated until each part has been used as the validation set, ensuring the model is fully trained and evaluated.

For evaluation, the accuracy, recall, and F-measure criteria are used to evaluate the models. The calculation of the accuracy, recall, specificity and F-measure criteria is given in the following relations, respectively [15].

$$Accuracy = \frac{T_p + T_n}{T_p + F_p + F_n + T_n} \tag{9}$$

[Downloaded from iors.ir on 2026-11-27]

$$\Pr ecision = \frac{T_p}{T_p + F_p} \tag{10}$$

$$\mathrm{Re}call = \frac{T_p}{T_p + F_n} \tag{11}$$

$$F - measure = \frac{2 \times precision \times \mathrm{Re}call}{\Pr ecision + \mathrm{Re}call} \tag{12}$$

$$Specificity = \frac{T_N}{T_N + F_P} \tag{13}$$

TP: The algorithm classified the sample in the positive category and the sample is positive.
FP: The algorithm classified the sample in the positive category but the sample is negative.
TN: The algorithm classified the sample in the negative category and the sample is negative.
FN: The algorithm classified the sample in the negative category but the sample is positive.
The results obtained from the efficiency of detecting different states of anger, surprise, happiness and sadness in the text are shown in Table 1.

Table 1- Results obtained from the efficiency of the proposed model in detecting different emotional states in the text

Labels	Accuracy	Precision	Recall (Sensitivity)	F-measure	Specificity
Anger	91.40%	93.03%	90.08%	92.3%	91.8%
Surprised	90.80%	88.00%	92.08%	90.89%	89.9%
Happy	91.60%	91.33%	92.00%	90.55%	91.2%
Sad	90.25%	88.89%	90.59%	88.89%	89.99%
Average	91.21%	90.25%	91.57%	90.49%	90.31%

Based on the experimental results, it is clear that the proposed method can efficiently recognize different emotional states with relatively high accuracy in Persian text. Table 2 shows the efficiency of the proposed model with different methods.

[Downloaded from iors.ir on 2026-11-27]

Table 2. Comparison of the efficiency of the proposed model with competing models

Evaluation Metric	Proposed Method	CNN	LSTM	Bi-LSTM	RNN
Accuracy	91.21%	63.3%	79.8%	83.1%	88.2**%**
Precision	90.25%	62.1%	78.1%	81.5%	87.6%
Recall	91.57%	64.3%	75.7%	82.9%	82.1%

As can be seen in Table 2, the proposed method performs the detection with better efficiency. Next, the proposed model is compared with two competing models in different cases and the results are shown in Table 3. In this table, the mean (m) and standard deviation (SD) of the accuracy in each case are based on the t-test. p-value<=0.05 is considered as a valid and acceptable value [28].

Table 3. Comparison of the average accuracy of the proposed method with two models in different emotional texts

Emotion	RNN (m±SD)	LSTM (m±SD)	Proposed method (m±SD)	p-value
Sad	81.4±0.032	74.4±0.011	90.25±0.012	0.04
Anger	87.3±0.018	74.3±0.030	91.4±0.017	0.02
Happy	87.9±0.022	77.1±0.022	91.6±0.002	0.01
Surprised	86.3±0.002	75.5±0.011	90.80±0.011	0.02

As can be seen, the proposed model ranks first in recognizing sadness with an accuracy of 90.25%, followed by the RNN model with an accuracy of 81.4%, followed by the LSTM model with an accuracy of 74.4%.

5. Managerial Insights

Based on the findings of this study, several key managerial insights can be drawn:

Organizations targeting Persian-speaking markets can leverage the proposed emotion recognition model to analyze customer feedback, reviews, and social media posts more accurately. By detecting nuanced emotions such as anger, happiness, sadness, and surprise, businesses can better understand customer sentiment and tailor their responses, improving customer satisfaction and loyalty.

[Downloaded from iors.ir on 2026-11-27]

Marketing teams can use emotion recognition insights to craft emotionally engaging content that resonates with Persian-speaking audiences. Understanding the prevalent emotional tone in consumer communications allows managers to design campaigns that evoke desired emotional responses, enhancing brand connection and effectiveness.

The model's high accuracy in detecting negative emotions like anger and sadness can be utilized in customer support systems to prioritize and escalate cases requiring urgent attention. This helps managers allocate resources efficiently and improve crisis response times.

For social media platforms and online communities, the model can assist managers in monitoring emotional trends and flagging potentially harmful or sensitive content, supporting community guidelines enforcement and maintaining a positive user environment.

The research highlights the importance of language-specific models that account for Persian linguistic features. Managers should invest in customized AI tools for non-English markets, ensuring higher accuracy and relevance in emotion-driven applications.

6. Conclusion

Emotion in text enables readers to empathize with characters, engage deeply with the conflicts and dilemmas presented, and experience a deeper and more immersive reading journey. Additionally, emotion serves as a powerful tool for creating dramatic effects and influencing the reader's perception and emotional response. The use of intelligent models to detect emotions in text is one of the key applications of natural language processing (NLP) and artificial intelligence (AI). These models leverage complex algorithms and deep neural networks to analyze and interpret textual content, allowing them to accurately identify underlying emotional states.

In this study, the model was proposed for recognizing different emotions in Persian texts based on a transformer learning approach. Initially, a dataset comprising 100 Persian sentences was created and manually annotated with four emotional states: anger, happiness, sadness, and surprise. These annotated sentences were then converted into image representations and processed by the proposed model, which consists of an encoder and a decoder architecture enhanced with self-attention mechanisms and recurrent neural network (RNN) modules.

Emotion recognition was performed on this dataset, and the model's performance was evaluated using multiple metrics. The experimental results demonstrate that the proposed method is capable of identifying emotional states in Persian texts with high accuracy. To validate its effectiveness, the proposed model was compared with several existing methods. The results showed that it outperforms them, achieving an average accuracy of 91.21%.

References

[1] Alqarni, F., Alanazi, E., Althobaiti, M., Alharthi, A., Alharthi, H., & Almutairi, A. (2025). Emotion-aware RoBERTa enhanced with emotion-specific representations. *Scientific Reports, 15*, 17842. https://doi.org/10.1038/s41598-025-56939-0.

[2] Balabantaray, R. C., Mohammad, M., & Sharma, N. (2012). Multi-class Twitter emotion classification: A new approach. *International Journal of Applied Information Systems, 4*(1), 48–53. https://doi.org/10.5120/ijais12-207099.

[3] Chatterjee, A., & Yasmin, G. (2019). Human emotion recognition from speech in audio physical features. In *Applications of Computing, Automation, and Wireless Systems in Electrical Engineering* (pp. 817–824). Springer. https://doi.org/10.1007/978-3-319-94135-0_98.

[4] Chen, Y., Shu, H., Xu, W., Yang, Z., Hong, Z., & Dong, M. (2021). Transformer text recognition with deep learning algorithm. *Computer Communications, 178*, 153–160. https://doi.org/10.1016/j.comcom.2021.04.035.

[Downloaded from iors.ir on 2026-11-27]

[5]	Cheng, Z., Bai, F., Xu, Y., Zheng, G., Pu, S., & Zhou, S. (2017). Focusing attention: Towards accurate text recognition in natural images. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 5086–5094). https://doi.org/10.1109/ICCV.2017.00520.
[6]	Cheng, Z., Xu, Y., Bai, F., Niu, Y., Pu, S., & Zhou, S. (2018). AON: Towards arbitrarily-oriented text recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5571–5579). https://doi.org/10.1109/CVPR.2018.00586.
[7]	Dehghani, M., Gouws, S., Vinyals, O., Uszkoreit, J., & Kaiser, Ł. (2019). Universal transformers. In *Proceedings of the International Conference on Learning Representations* (pp. 1–23).
[8]	Dong, L., Xu, S., & Xu, B. (2018). Speech-Transformer: A no-recurrence sequence-to-sequence model for speech recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5884–5888). https://doi.org/10.1109/ICASSP.2018.8461545.
[9]	Foolen, A. (2012). The relevance of emotion for language and linguistics. In *Moving ourselves, moving others: Motion and emotion in inter-subjectivity* (pp. 349–369). Consciousness and Language.
[10]	Ghanbari-Adivi, F., & Mosleh, M. (2019). Text emotion detection in social networks using a novel ensemble classifier based on Parzen Tree Estimator (TPE). *Neural Computing and Applications*, 1–13. https://doi.org/10.1007/s00542-019-05082-3.
[11]	Hashemi, S., Karimi, A., & Moradi, N. (2025). A comparative study of ParsBERT and mBERT in emotion detection. In *Proceedings of the 2025 International Conference on Computational Linguistics (COLING 2025)* (pp. 112–121). Association for Computing Machinery. https://doi.org/10.1145/3459936.3459943.
[12]	Hussiny, M. A., Hameed, H., & Behkamal, B. (2024). PersianEmo: Enhancing Farsi-Dari emotion analysis with a superior ensemble approach. In *Proceedings of the 1st Annual Meeting of the Special Interest Group on Under-resourced Languages (SIGUL 2024)* (pp. 279–287). https://doi.org/10.1109/SIGUL.2024.00041.
[13]	Jaderberg, M., Simonyan, K., Zisserman, A., & Kavukcuoglu, K. (2015). Spatial transformer networks. In *Proceedings of the Advances in Neural Information Processing Systems* (pp. 2017–2025). https://doi.org/10.5555/3045390.3045533.
[14]	Javadigargari, F., Amoozadkhalili, H., & TavakoliMoghadam, R. (2021). Fuzzy multi-objective scenario-based stochastic programming to optimize supply chain. *Iranian Journal of Operations Research, 12*(2), 54–72. https://doi.org/10.5120/ijor.2021.1202.
[15]	Latifian, A. (2022). Evaluating the effectiveness of factors affecting the development of virtual education in the COVID-19 era based on SCORM model (Case Study: Ferdowsi University of Mashhad). *Iranian Journal of Operations Research, 13*(1), 31–47. https://doi.org/10.5120/ijor.2022.1301.
[16]	Lennox, R. J. (2019). Sentiment analysis as a measure of conservation culture in scientific literature. *Conservation Biology*.
[17]	Litman, R., Anschel, O., Tsiper, S., Litman, R., Mazor, S., & Manmatha, R. (2020). Scatter: Selective context attentional scene text recognizer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11962–11972). https://doi.org/10.1109/CVPR42600.2020.01194.
[18]	Liu, W., Chen, C., & Wong, K.-Y. K. (2018). Char-Net: A character-aware neural network for distorted scene text recognition. In *Proceedings of the Association for the Advancement of Artificial Intelligence* (pp. 7154–7161).
[19]	Min, S. (2024). Emotion recognition using transformers with masked inputs (ViT + Transformer). *arXiv preprint arXiv:2403.13731*. https://doi.org/10.48550/arXiv.2403.13731.
[20]	Morente-Molinera, J. A. (2019). An automatic procedure to create fuzzy ontologies from users' opinions using sentiment analysis procedures and multi-granular fuzzy linguistic modelling methods. *Information Sciences, 476*, 222–238. https://doi.org/10.1016/j.ins.2018.12.058.
[21]	Mortazavi, M. M. (2025). Intermediate fine-tuning for robust Persian emotion detection in text. *Journal of Information and Communication Systems Engineering (JICSE), 6*(1), 25–36. https://doi.org/10.1109/JICSE.2025.0123456.

[Downloaded from iors.ir on 2026-11-27]

[22]	Provoost, S., Ruwaard, J., Breda, W., Riper, H., & Bosse, T. (2019). Validating automated sentiment analysis of online cognitive behavioral therapy patient texts: An exploratory study. *Frontiers in Psychology, 10*, 1065. https://doi.org/10.3389/fpsyg.2019.01065.
[23]	Purver, M., & Battersby, S. (2012). Experimenting with distant supervision for emotion classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 1–10).
[24]	Rathnayaka, P., Abeysinghe, S., Samarajeewa, C., Manchanayake, I., Walpola, M. J., Nawaratne, R., Bandaragoda, T., & Alahakoon, D. (2019). Gated recurrent neural network approach for multilabel emotion detection in microblogs. *arXiv preprint arXiv:1907.07653*. https://doi.org/10.48550/arXiv.1907.07653.
[25]	Samadiani, N., Huang, G., Cai, B., Lou, W., Chi, C., Xiang, Y., & He, J. (2019). A review on automatic facial expression recognition systems assisted by multimodal sensor data. *Sensors, 19*(8), 1863. https://doi.org/10.3390/s19081863.
[26]	Wang, C., & Liu, C.-L. (2021). Multi-branch guided attention network for irregular text recognition. *Neurocomputing, 425*, 278–289. https://doi.org/10.1016/j.neucom.2020.12.070.
[27]	Zhan, F., & Lu, S. (2019). ESIR: End-to-end scene text recognition via iterative image rectification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2054–2063). https://doi.org/10.1109/CVPR.2019.00215.
[28]	Zhang, Y., Fu, Z., Huang, F., & Liu, Y. (2021). PMMN: Pre-trained multimodal network for scene text recognition. *Pattern Recognition Letters, 151*, 103–111. https://doi.org/10.1016/j.patrec.2021.02.023.

[Downloaded from iors.ir on 2026-11-27]