6th International Conference on Web Research
June 11, 2020; Tehran, IRAN
ICWR2020

Bidirectional LSTM Deep Model for Online Doctor Reviews Polarity Detection

Mohammad Ehsan Basiri*, Reza Salehi Chegeni, Aria Naseri Karimvand, Shahla Nemati
Department of Computer Engineering, Shahrekord University, Shahrekord, Iran.
basiri@sku.ac.ir, reza.sch@yahoo.com, aria.nk@yahoo.com, s.nemati@sku.ac.ir

***Abstract*—Online medical reviews contain patients' subjective evaluations and reflect their satisfaction with the treatment process and doctors. Mining and analysis of sentiment expressed in these medical data may be vital for different applications including adverse drug effects detection, doctor recommendation, and healthcare quality assessment. Nevertheless, medical sentiment analysis is a challenging and complex task because patients who write the reviews are usually non-professional users and tend to use informal language. The problem is more challenging in the Persian language due to its resource scarcity and complex structure. In this study, we introduce PODOR, a Persian dataset of online doctor reviews extracted from social web. Also, we propose a deep model based on the bidirectional long short-term memory for polarity detection of PODOR reviews. To show the effectiveness and suitability of the proposed model, we compared the model with six traditional supervised machine learning methods and three deep models. Preliminary comparative results indicated that our model outperformed traditional methods by 8% and 7%, and deep models by 2% and 3% in terms of accuracy and f1-measure.**

Keywords—Online Doctor Reviews; Patients Opinion; Sentiment Analysis; Medical Language Processing; Deep Learning.

I. INTRODUCTION

With the emergence of fast-growing social media, patients have now the opportunity to express their opinion regarding doctors in online platforms such as healthgrades[1] and WebMD[2]. These platforms provide objective information such as details of specialists near the user, type of insurance they accept, and distance from you, as well as subjective information such as percentage of patients that recommend the doctor and patient satisfaction reviews [1]. Mining and analysis of subjective information provided in user reviews can help other patients who are interested to know how well a doctor provided treatment or care. Moreover, doctor recommendation systems may use such information to help patients choose a doctor to treat a disease [2].

Treating subjectivity in textual information is focused in sentiment analysis (SA) which is a subfield of natural language processing (NLP) [3]. The main goal of SA is to extract and detect sentiment, opinion, attitude, and appraisal from free-form text [4]. SA is a relatively new field of study, though, it is an extensively studied domain in processing unstructured textual data published in social media [5]. Researchers in fields of SA often concentrate on processing reviews in domains such as marketing, political, and tourism and very few research studies addressed reviews in medical social media [6].

Medical sentiment has some features making it an interesting source of information for SA research [1]. Nevertheless, text information derived from websites focused on healthcare and medical domain are more complicated with the medical terminology [7]. Another difficulty in texts in medical domain is the indirect expression of sentiment [7]. Moreover, high lexical diversity resulted from the joint use of informal language and specific terminology makes medical SA more challenging [8].

Two types of subjective medical patients-written text exist in social media; Drug reviews and Online Doctor Reviews (ODRs). Drug reviews are anonymously written texts describing different aspects of drug usage such as side effects, value, and effectiveness [1]. ODRs are usually written by patients that intend to express their opinion about the doctor or treatment process they experienced. These reviews are used as the very first step for patients to finding a new doctor. There are some similarities between drug reviews and ODRs. For example, both are usually written by non-professional users and both describe user experience with medical settings [8]. However, it has been shown that ODRs contain less technical terminology and are shorter [8].

In the current study, we investigate ODRs written in the Persian language because among Persian websites, there exist more ODR websites on which users may express their opinion in comparison to drug reviews. Moreover, the results of this analysis may be used in in doctor recommender systems [2]. To the best of our knowledge, previous studies have not addressed Persian ODR sentiment analysis. Persian SA has several difficulties such as resource scarcity, grammatical and structural complexities of Persian language, and problems with informal language used in Persian social

[1] https://www.healthgrades.com
[2] https://www.webmd.com

media [9]–[11]. Most studies in sentiment analysis of English ODRs employ traditional machine learning (ML) methods such as support vector machine (SVM), decision trees and naïve Bayes (NB) [12]. These methods usually exploit n-gram and lexical features and have the limitation of ignoring the semantic information embodied in the text [12]. Also, feature engineering that is a vital step in traditional ML methods is a time- and labor-intensive task [13].

In order to address the above mentioned problems, in this study a deep learning model is proposed for sentiment analysis of Persian ODRs. In the proposed model bidirectional long short-term memory (BiLSTM) is employed to utilize sequential dependencies and contextual information in the text. Specifically, by processing reviews from both directions, BiLSTM is able to use both past and future context in the reviews. To assess the performance of the proposed method, we crawled a dataset of Persian ODRs, PODOR, from www.nobat.ir website. Reviews in this website are rated by users in a 5-star scale. We mapped the 5-star ratings into polarity labels in a same way described in [6]. In summary, this study contributes in the following three folds:

- We introduce the first Persian ODR dataset, PODOR which contains reviews and their polarity.
- We propose a deep neural network model that considers contextual information in the reviews.
- We compare our proposed model with six traditional machine learning and three deep neural network models.

The paper is structured in the following way: Section II briefly presents a literature review of SA studies in the medical domain. In Section III, the proposed model for sentiment analysis of ODRs is described. Sections IV demonstrates the results obtained and discusses them. Finally, conclusions and future work are presented in Section V.

II. Literature Review

This section is divided into two parts; a brief overview of Persian sentiment analysis methods and medical sentiment analysis.

A. Sentimetn analysis in the Persian language

Sentiment analysis research in the Persian language has been started from 2012 [14] and can be divided into three categories: lexicon-based, machine learning (ML), and hybrid methods. Lexicon-based methods have some advantages such as simplicity and less computational complexity whereas ML methods are more accurate and less domain dependent [15]. Hybrid methods were proposed to utilize the advantages of both lexicon-based and ML-based methods.

Because the Persian language is considered as a resource limited language [16], some researchers developed manually created lexicons. For example, Basiri and Kabiri [17] introduce CNRC and Adjective lexicons and showed that direct translation of lexicons is not sufficient for lexicon-based SA in the Persian language. Dashtipour et al. [16] proposed PerSent which is a list of 1500 words, their part-of-speech (POS) tags, and sentiment polarity scores. Recently, Basiri and Kabiri [11] introduce PerLex and evaluated its performance on a large dataset of Persian reviews. They showed that PerLex achieves a higher accuracy than existing Persian lexicons for sentiment analysis. There are also some other lexicons for Persian SA that are not listed here due to space limitation.

Several ML techniques were employed for Persian SA in the literature. For example, in [18], SVM is used for classification of Persian reviews and in [19], Naïve Bayes is used along with a feature selection method for polarity detection. Recently, deep neural networks have also been exploited for Persian SA. For example, Roshanfekr et al. [20] introduced a dataset of electronic products reviews and evaluated some deep learning methods. Zobeidi et al. [21] proposed a convolutional neural network (CNN) for feature selection in Persian SA and Dastgheib et al. [22] utilized a combination of structural correspondence learning (SCL) and CNN for Persian SA.

Some researchers tried to improve Persian SA using different approaches such as improving score aggregation [6], [23], utilizing feature selection [24], [25], and considering targets [26]. In a different way, some researchers proposed hybrid methods. For example, Dashtipour et al. [27] proposed a rule-based method using dependency grammar and deep learning for polarity detection in the Persian language. They compared their method with both traditional and deep learning methods and reported an improvement over both methods. Basiri and Kabiri [28] proposed a hybrid of ML and lexicon-based methods for Persian SA. They utilized three lexicons for feature extraction and combined these features with common NLP features.

B. Medical sentiment analysis

Early studies on medical content analysis showed that different medical social media text provide different types of medical and affective information [29]. It has been also shown that medical and healthcare related texts contain objective and subjective information and SA can be employed to assess positive or negative medical outcomes and patients' conditions [1]. Existing research on medical SA may be categorized into three groups with respect to the target they concern [1]. The first group of research studies consider the sentiment as a reflection of patients' health status. The second and third groups concern the sentiment as judgement of medical conditions and outcome of a treatment, respectively.

With respect to their underlying classification method, existing methods for medical SA can be categorized into lexicon-based and ML-based methods [1]. Liu and Lee [7] proposed a feature extraction method for drug review SA and introduce a medical domain sentiment lexicon. Their method considers position embeddings to represent position encoding. Asghar et al. [30] generate SentiHealth, a domain specific lexicon for health-related users' SA. Another lexicon-based method proposed by Zhang et al. [2] in a medical recommendation system. They utilized SentiWordNet 3.0 [31] and negation rules to calculate sentiment score of reviews.

However, most studies exploited ML-based techniques to classify the sentiment in the medical resource. For example, Jiménez-Zafra et al. [8] introduce two corpus of Spanish medical reviews and applied SVM and compared the results with a lexicon-based approach. Mondal et al. [32] applied different ML methods on the lexical features extracted using existing medical lexicons. Deep learning methods have been also used in medical SA. For example, Shi et al. proposed a multi-aspect attention mechanism for document-level SA of reviews written by medical experts. Edara et al. [33] proposed a long short-term memory (LSTM) for SA of cancer medical records including tweets medical abstracts. Sharma et al. [34] used CNNs for predicting review rating from its text.

III. PROPOSED MODEL

Fig. 1 shows the overall view of the proposed deep model for online doctor review sentiment analysis.

A. Input and embedding

As shown in the figure, the input to the model is a review from PODOR dataset. Before a review can be presented to the model, the review must be first encoded so that each word in the review is represented by a unique integer. After analyzing the PODOR dataset, 50 is considered as the length of each input sequence. Because the lengths of reviews in the PODOR are not equal, to make all the vectors of same length, zero padding is used. The first layer of the network is an embedding layer that map the word indexes to their dense vector representations. In this study, we considered the embedding dimension as 20.

B. Bidirectional LSTM

To model sequence data, recurrent neural networks (RNNs) were proposed [35]. RNNs are not suitable in cases where the sequence is long enough. For such cases, LSTMs and gated recurrent networks (GRUs) are good solutions. Both LSTM and GRU are proposed to address the vanishing gradient problem of original RNNs [36]. The main difference between LSTM and GRU is their gates. We preferred to use LSTM in the current study as it has shown promising results in similar problems, recently [36], [37][37]. LSTM models address the short-term memory limitation of RNNs using the cell state, and their different types of gates [12]. The memory cell, c_t, in

LSTM is used to preserve the state over the time. Input gate, i_t, forget gate, f_t, and output gate, o_t, are used for information flow regulation. These gates are sigmoid activation functions, hence output only values in range [0,1]. Equations (1) to (3) are used in the gates:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (1)$$

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (2)$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (3)$$

where, σ is sigmoid function, w_x, h_{t-1}, x_t, and b_x are weight for gate x, output of the previous LSTM cell, input at current step, and bias for gate x, respectively. Input gate is used to decide what new information should be stored in the cell, forget gate decides which information should be thrown away the cell, and output gate provide the activation to the final output at step t.

Equations (4) to (6) are used to show cell state, s_t, candidate cell state, $\tilde{s}$, and final output, h_t.

$$\tilde{s}_t = \tanh(w_c[h_{t-1}, x_t] + b_c) \quad (4)$$

$$s_t = f_t \odot s_{t-1} + i_t \odot \tilde{s}_t \quad (5)$$

$$h_t = o_t \odot \tanh(s_t) \quad (6)$$

where $\odot$ represents the element wise multiplication of the vectors. The block diagram of the LSTM cell is shown in Fig. 2.

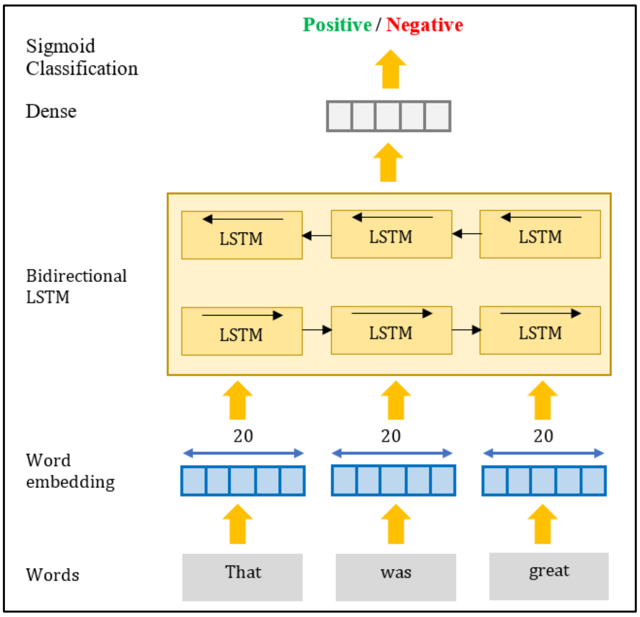

Fig. 1. The overall view of the proposed deep model.

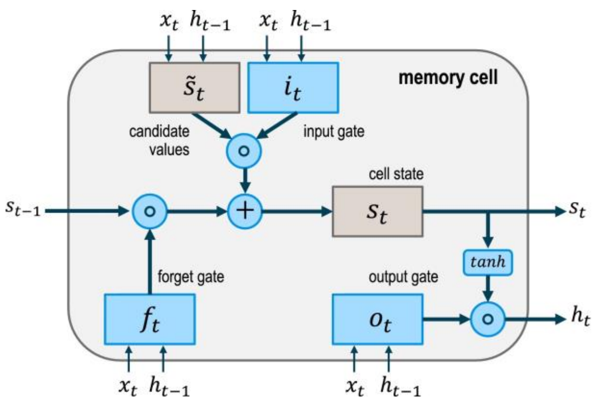

Fig. 2. The block diagram of an LSTM cell adapted from [38].

BiLSTM cells contain a forward LSTM with output sequence $\vec{h}$, and a backward LSTM layer with output $\overleftarrow{h}$. In BiLSTM cells, $\vec{h}$ can be calculated using time steps in a normal direction from time $T-n$ to $T-1$ and $\overleftarrow{h}$ can be calculated in a reverse order from $T-1$ to $T-n$. This lets the network processes the information flow in both directions and improves model performance. In our ODR classification problem, we employed BiLSTM as all time-steps of the input sequence are available in advance. The output of the BiLSTM layer is passed to a dense layer for final classification.

IV. Experiments and Results

In this section, we first present the specification of the dataset we introduce and then, performance measures used to evaluate different methods are described. Finally, the results obtained using the proposed and other methods will be discussed.

A. Dataset

Because there is no public dataset for Persian ODR, we crawled PODOR containing 700 reviews about different doctors from www.nobat.ir published from 2018 to 2019. These reviews are rated by users who posted the review in a 5-star scale. In order to convert the ratings to polarity labels, we divided the [0,5] interval into two equal parts and consider the reviews belong to the lower and higher parts as the negative and positive classes, respectively. More details of the PODOR dataset is shown in Table I. Fig. 3 and Fig. 4 show the word cloud of PODOR dataset in Persian and English, respectively.

B. Performance measures

Four performance measures are used in the experiments to assess the performance of models according to (7) to (10):

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$F1-score = \frac{2\times Pr\times Re}{Pr+Re} \quad (9)$$

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (10)$$

where *TP*, *TN*, *FP*, and *FN* are true positive, true negative, false positive, and false negative, respectively.

C. Experimental setup

All experiments were carried out on a machine with two Intel(R) Xeon(R) 2.00GHz CPUs with 6MB cache, 13GB RAM, and a Tesla K80 GPU with 12GB GDDR5 VRAM. All implementations were coded in Python 3.6 in Google Colab environment and Keras library [39] was used for implementing deep models.

To show the utility of the proposed deep model, its performance is compared to those obtained using six traditional machine learning methods, namely, artificial neural network (ANN), support vector classifier (SVC), K-nearest neighbor (KNN), decision tree (DT), Naïve Bayes (NB), random forest (RF), and three deep models, namely, LSTM [37], 3CNN [40], and 3CNN-LSTM [41]. Parameter settings of traditional learning models are shown in Table II.

TABLE I. Specification of PODOR dataset. The second and third columns show average word count before and after stop-word removal.

Class	***# of reviews***	***avg word before stop-word removal***	***avg word after stop-word removal***	***avg Sentence***
Negative	427	14.07	8.87	1.29
Positive	273	29.27	17.33	3.31
Total	700	21.67	13.1	2.3

Fig. 3. Word cloud of the PODOR dataset in Persian.

Fig. 4. Word cloud of the PODOR dataset in English.

D. Results

The accuracy of all 10 models are compared in Fig. 5. Moreover, the receiver operating characteristic (ROC) curves of traditional and deep methods are compared in Fig. 6 and Fig. 7, respectively.

As can be seen in the figures, all deep models outperform traditional ML methods. This may be due to the fact that deep models extract more meaningful features from the ODR reviews. Moreover, the proposed BiLSTM model outperforms LSTM, 3CNN, and 3CNN-LSTM models. This shows the effect of using contextual information in both forward and reverse directions in the proposed BiLSTM model. In fact, as there may be long dependencies between words in an ODR, BiLSTM are more suitable in comparison to CNN models. This may be the cause for obtaining better results by our model.

To compare the methods in more details, Table III shows the precision, recall, and F1-score of all 10 methods for the positive and negative classes separately. As shown in Table III, the best results obtained using the proposed BiLSTM model for all measures.

Time comparison of models are shown in Table IV. As show in the table, all traditional supervised methods have lower training and test times in comparison to those for deep models. This is the advantage of using traditional models.

The proposed model has a higher training time in comparison to other deep models. This is due to the use of bidirectional LSTM cells in the proposed model. However, the test time for all deep models are equal.

V. Conclusion

Online doctor reviews are informative source of medical knowledge that may be used by patients to decide about selecting doctors and specialists. These valuable reviews are written in an informal way by non-professional users and detecting their sentiment may help other patients find related reviews. Nevertheless, due to higher lexical diversity and use of indirect expression of sentiment, medical text information is more complicated in comparison to product related reviews. Also, for the lack of such medical resources for the Persian language resulted to lack of research on this domain. To address these problems, in the current study, first, we introduce PODOR, a dataset of online doctor reviews extracted from www.nobat.ir website. Then, we proposed a deep learning model for sentiment analysis of PODOR reviews using bidirectional long short-term memory. This proposed model is able to use both past and future context in the reviews. Results of comparing the proposed model with six traditional machine learning algorithms and three deep models showed the superiority of the proposed model in processing the medical reviews. For the future work, the authors plan to extend the PODOR dataset with more doctor reviews and comparing the performance of the model on similar datasets in the English language. Another line of research may be enhancing the proposed model by considering attention mechanism on the top of the LSTM layer of the proposed model.

TABLE II. Parameter settings of traditional ML methods.

Classifier	***Parameters***
ANN	hidden_layer_sizes = (100,), activation = relu, solver = adam, alpha = 0.0001, learning_rate = 0.001, max_iter = 200
SVC	loss = squared hinge, tolerance = 0.0001, regularization parameter = 0.1, max_iter = 1000
KNN	neighbors = 5, weights = uniform, leaf_size = 30, metric = minkowski, distance = euclidean
NB	multinomial model, alpha = 1.0, fit_prior = True
DT	criterion = gini, splitter = best, min_samples_split = 2, min_samples_leaf = 1
RF	n_estimators = 100, criterion = gini, min_samples_split = 2, min_samples_leaf = 1

TABLE III. Comparison of six traditional and three deep models with the proposed BiLSTM model.

Method	**Class**	***Precision***	***Recall***	***F1-score***
ANN [42]	Negative	0.47	0.40	0.43
	Positive	0.63	0.69	0.66
	avg	0.56	0.57	0.57
SVC [42]	Negative	0.69	0.44	0.54
	Positive	0.69	0.87	0.77
	avg	0.69	0.69	0.68
KNN [42]	Negative	0.65	0.56	0.60
	Positive	0.73	0.80	0.76
	avg	0.70	0.70	0.70
DT [42]	Negative	0.60	0.58	0.59
	Positive	0.72	0.73	0.73
	avg	0.67	0.67	0.67
NB [42]	Negative	0.51	0.89	0.65
	Positive	0.85	0.41	0.55
	avg	0.71	0.61	0.59
RF [42]	Negative	0.70	0.61	0.65
	Positive	0.76	0.82	0.79
	avg	0.73	0.74	0.73
LSTM [37]	Negative	0.70	0.67	0.68
	Positive	0.78	0.81	0.79
	avg	0.75	0.75	0.75
CNN-LSTM [41]	Negative	0.76	0.67	0.71
	Positive	0.79	0.86	0.82
	avg	0.78	0.78	0.78
3CNN [40]	Negative	0.77	0.70	0.73
	Positive	0.81	0.86	0.83
	avg	0.79	0.79	0.79
BiLSTM (This study)	Negative	0.76	0.79	**0.78**
	Positive	**0.85**	**0.83**	**0.84**
	avg	**0.82**	**0.81**	**0.81**

TABLE IV. Comparison of six traditional and three deep models with the proposed BiLSTM model.

Method	***Train time (S)***	***Test time (S)***
ANN [42]	3	<1
SVC [42]	3	<1
KNN [42]	2	<1
DT [42]	2	<1
NB [42]	2	<1
RF [42]	2	<1
LSTM [37]	99	2
CNN-LSTM [41]	40	2
3CNN [40]	79	2
BiLSTM (This study)	113	2

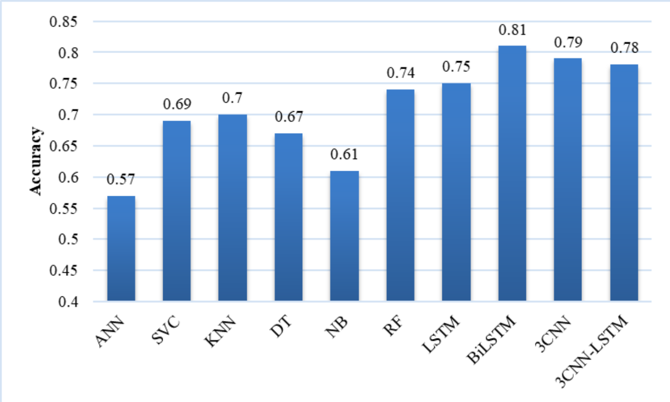

Fig. 5. Comparison of the accuracy obtained using six traditional and three deep models with the proposed BiLSTM model.

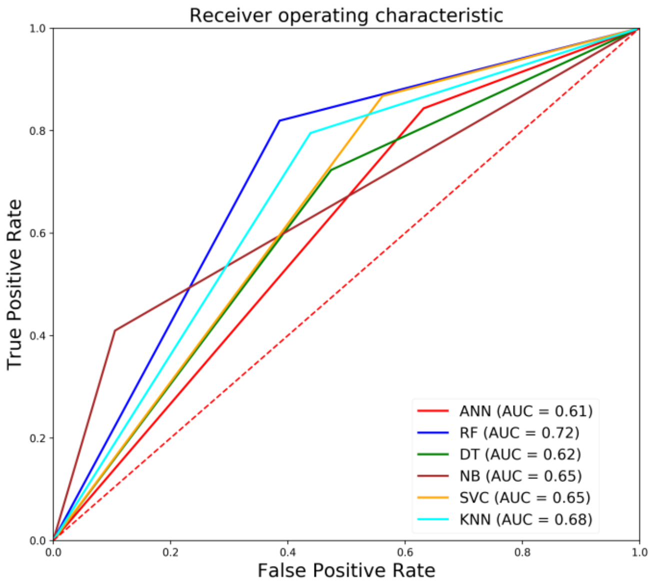

Fig. 6. Comparison of ROC curves for traditional ML methods.

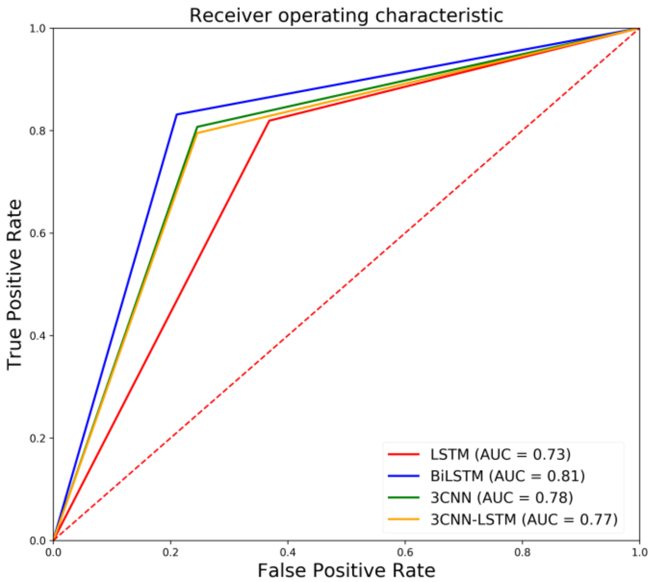

Fig. 7. Comparison of ROC curves for deep methods.

REFERENCES

[1] K. Denecke and Y. Deng, "Sentiment analysis in medical settings: New opportunities and challenges," Artificial intelligence in medicine, vol. 64, no. 1, pp. 17–27, 2015.

[2] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, "iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization," Future Generation Computer Systems, vol. 66, pp. 30–35, 2017.

[3] E. Cambria, "Affective computing and sentiment analysis," IEEE Intelligent Systems, vol. 31, no. 3, pp. 102–107, 2016.

[4] B. Liu, Sentiment analysis: Mining opinions, sentiments, and emotions, Cambridge University Press, 2015.

[5] S. Nemati, "OWA operators for the fusion of social networks' comments with audio-visual content," in Proc. 5th International Conference on Web Research (ICWR), 2019, pp. 90–95.

[6] M. E. Basiri, A. Kabiri, M. Abdar, W. K. Mashwani, N. Y. Yen, and J. C. Hung, "The effect of aggregation methods on sentiment classification in Persian reviews," Enterprise Information Systems, pp. 1–28, 2019.

[7] S. Liu and I. Lee, "Extracting features with medical sentiment lexicon and position encoding for drug reviews," Health information science and systems, vol. 7, no. 1, pp. 11-21, 2019.

[8] S. M. Jiménez-Zafra, M. T. Martin-Valdivia, M. D. Molina-González, and L. A. Ureña-López, "How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain," Artificial intelligence in medicine, vol. 93, pp. 50–57, 2019.

[9] F. Amiri, S. Scerri, M. H. Khodashahi, F. Iais, and S. Augustin, "Lexicon-based Sentiment Analysis for Persian Text," pp. 9–16, 2015.

[10] M. E. Basiri and A. Kabiri, "Sentence-level sentiment analysis in Proc. 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA), pp. 84–89, Apr. 2017.

[11] M. E. Basiri and A. Kabiri, "Words Are Important: Improving Sentiment Analysis in the Persian Language by Lexicon Refining," ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), vol. 17, no. 4, pp. 26-31, 2018.

[12] A. M. Shah, X. Yan, S. A. A. Shah, and G. Mamirkulova, "Mining patient opinion to evaluate the service quality in healthcare: a deep-learning approach," Journal of Ambient Intelligence and Humanized Computing, pp. 1–18, 2019.

[13] S. Nemati, R. Rohani, M. E. Basiri, M. Abdar, N. Y. Yen, and V. Makarenkov, "A Hybrid Latent Space Data Fusion Method for Multimodal Emotion Recognition," IEEE Access, vol. 7, pp. 172948–172964, 2019.

[14] M. Shams, A. Shakery, and H. Faili, "A non-parametric LDA-based induction method for sentiment analysis," in Proc. The 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP 2012), 2012, pp. 216–221.

[15] M. E. Basiri, N. Ghasem-Aghaee, and A. R. Naghsh-nilchi, "Lexicon-based Sentiment Analysis in Persian," Current and Future Developments in Artificial Intelligence, pp. 154-184, 2017.

[16] K. Dashtipour, A. Hussain, Q. Zhou, A. Gelbukh, A. Y. A. Hawalah, and E. Cambria, "PerSent: A Freely Available Persian Sentiment Lexicon," In Proc. International Conference on Brain Inspired Cognitive Systems, 2016, pp. 310–320.

[17] M. E. Basiri and A. Kabiri, "Translation is not enough: comparing lexicon-based methods for sentiment analysis in Persian.," in Proc. 2017 International Symposium on Computer Science and Software Engineering Conference (CSSE) IEEE, 2017, pp. 36-41.

[18] M. S. Hajmohammadi and R. Ibrahim, "A SVM-based method for sentiment analysis in Persian language," in Proc. International Conference on Graphic and Image Processing (ICGIP), 2013, pp. 876-883.

[19] S. M. Bagheri A., "Persian sentiment analyzer: A framework based on a novel feature selection method," International Journal of Artificial Intelligence, vol. 12, no. 2, pp. 115–129, 2014.

[20] B. Roshanfekr, S. Khadivi, and M. Rahmati, “Sentiment analysis using deep learning on Persian texts,” in Proc. 2017 Iranian Conference on Electrical Engineering (ICEE), 2017, pp. 1503–1508.

[21] S. Zobeidi, M. Naderan, and S. E. Alavi, “Opinion mining in Persian language using a hybrid feature extraction approach based on convolutional neural network,” Multimedia Tools and Applications, vol. 78, no. 22, pp. 32357–32378, 2019.

[22] M. B. Dastgheib, S. Koleini, and F. Rasti, “The application of Deep Learning in Persian Documents Sentiment Analysis,” International Journal of Information Science and Management (IJISM), vol. 18, no. 1, pp. 1–15, 2020.

[23] M. E. Basiri and A. Kabiri, “Uninorm operators for sentence-level score aggregation in sentiment analysis,” in Proc. 4th International Conference on Web Research (ICWR), 2018, pp. 97–102.

[24] M. Saraee and A. Bagheri, “Feature Selection Methods in Persian Sentiment Analysis,” In Proc. International Conference on Application of Natural Language to Information Systems, 2013, pp. 303-308.

[25] A. Bagheri, M. Saraee, and F. de Jong, “Sentiment classification in Persian: Introducing a mutual information-based method for feature selection,” in Proc. 21st Iranian Conference on Electrical Engineering (ICEE), 2013, pp. 1–6.

[26] M. E. Basiri, M. Abdar, A. Kabiri, S. Nemati, X. Zhou, F. Allahbakhshi, and N. Y. Yen., “Improving Sentiment Polarity Detection Through Target Identification.” IEEE Transactions on Computational Social Systems, vol. 7, no. 1, pp. 113-128, 2019.

[27] K. Dashtipour, M. Gogate, J. Li, F. Jiang, B. Kong, and A. Hussain, “A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks,” Neurocomputing, vol. 380, pp. 1–10, 2020.

[28] M. E. Basiri and A. Kabiri, “HOMPer: A new hybrid system for opinion mining in the Persian language,” Journal of Information Science, vol. 46, no. 1, pp. 101-117, 2019.

[29] K. Denecke and W. Nejdl, “How valuable is medical social media data? Content analysis of the medical web,” Information Sciences, vol. 179, no. 12, pp. 1870–1880, 2009.

[30] M. Z. Asghar, S. Ahmad, M. Qasim, S. R. Zahra, and F. M. Kundi, “SentiHealth: creating health-related sentiment lexicon using hybrid approach,” SpringerPlus, vol. 5, no. 1, pp. 1139, 2016.

[31] S. Baccianella, A. Esuli, and F. Sebastiani, “Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining” in Proc. LREC, vol. 10, 2010, pp. 2200–2204.

[32] A. Mondal, E. Cambria, D. Das, A. Hussain, and S. Bandyopadhyay, “Relation extraction of medical concepts using categorization and sentiment analysis,” Cognitive Computation, vol. 10, no. 4, pp. 670–685, 2018.

[33] D. C. Edara, L. P. Vanukuri, V. Sistla, and V. K. K. Kolli, “Sentiment analysis and text categorization of cancer medical records with LSTM,” Journal of Ambient Intelligence and Humanized Computing, pp. 1–17, 2019.

[34] R. D. Sharma, S. Tripathi, S. K. Sahu, S. Mittal, and A. Anand, “Predicting online doctor ratings from user reviews using convolutional neural networks,” International Journal of Machine Learning and Computing, vol. 6, no. 2, pp. 149-161, 2016.

[35] T. Mikolov, S. Kombrink, L. Burget, J. Černocký, S. Khudanpur, “Extensions of recurrent neural network language model. ” In Proc. 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2011, pp. 5528-5531

[36] G. Rao, W. Huang, Z. Feng, and Q. Cong, “LSTM with sentence representations for document-level sentiment classification,” Neurocomputing, vol. 308, no. 1, pp. 49–57, 2018.

[37] J. Nowak, A. Taspinar, and R. Scherer, “LSTM recurrent neural networks for short text and sentiment classification,” in Proc. International Conference on Artificial Intelligence and Soft Computing, 2017, pp. 553–562.

[38] T. Fischer and C. Krauss, “Deep learning with long short-term memory networks for financial market predictions,” European Journal of Operational Research, vol. 270, no. 2, pp. 654–669, 2018.

[39] F. Chollet and others, “Keras.” GitHub, 2015.

[40] Y. Kim, “Convolutional neural networks for sentence classification,” arXiv preprint arXiv:1408.5882, 2014.

[41] X. Wang, W. Jiang, and Z. Luo, “Combination of convolutional and recurrent neural network for sentiment analysis of short texts,” in Proc. COLING 2016, the 26th international conference on computational linguistics: Technical papers, 2016, pp. 2428–2437.

[42] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of machine learning research, vol. 12, no. Oct, pp. 2825–2830, 2011.