



Deep learning based analysis of sentiment dynamics in online cancer community forums: An experience

Athira Balakrishnan 

Sumam Mary Idicula

Cochin University of Science and Technology, India

Josette Jones

Indiana University, USA

Abstract

Online health communities (OHC) provide various opportunities for patients with chronic or life-threatening illnesses, especially for cancer patients and survivors. A better understanding of the sentiment dynamics of patients in OHCs can help in the precise formulation of the needs during their treatment. The current study investigated the sentiment dynamics in patients' narratives in a Breast Cancer community group (Breastcancer.org) to identify the changes in emotions, thoughts, stress, and coping mechanisms while undergoing treatment options, particularly chemotherapy, radiation, and surgery. Sentiment dynamics of users' posts was performed using a deep learning model. A sentiment change analysis was performed to measure change in the satisfaction level of the users. The deep learning model BiLSTM with sentiment embedding features provided a better F1-score of 91.9%. Sentiment dynamics can assess the difference in satisfaction level the users acquire by interacting with other users in the forum. A comparison of the proposed model with existing models revealed the effectiveness of this methodology.

Keywords

breast cancer community, deep-learning, machine learning, sentiment analysis, sentiment dynamics

Corresponding author:

Athira Balakrishnan, Department of Computer Science, Cochin University of Science and Technology, Cochin, Kerala 682022, India.

Email: athiramalu@gmail.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Introduction

In OHCs, people form community groups, talk about personal health conditions, and seek support from peers. Patients and their families' caregivers undergo physical and mental consequences, such as stress, anxiety, and depression due to cancer and cancer care.¹ A study suggests that online community networks are useful in reducing stress, increasing self-efficacy and improving quality of life.² Although health-related sites offer a wide range of information, gaining knowledge from similar experienced people add significant value. OHCs mainly center on three types of support: informational support, emotional support, and companionship.³

While early detection and timely and efficient care for breast cancer is growing, it is also essential to learn how to handle the condition and preserve everyday life quality. Although many studies have been performed on cancer survival, not much research work is reported about the "experience" of living with cancer. The factors associated with unmet needs in post-treatment cancer survivors are fear of recurrence, up to date information, and more reduced quality of life.⁴ Seeking and providing additional support with similar patients is one of the critical needs for patients living with cancer.

Sentiment analysis (SA) aims to analyze large amounts of information shared through online support groups and to detect the discourse structure and characteristics. Some SA applications include identifying prominent group participants, facilitating the search for information in OHCs, and considering the social relationships in OHCs.⁵ Most of the literature in OHCs showed the studies were focused on finding the community's dynamism or finding the influential members or factors in the community, but rarely investigated their current needs and future needs.⁶⁻⁸ The current study focused on sentiment dynamics of patients in a breast cancer community group. Emotional dynamics are highest during the treatment phase and hence the patient's emotions were analyzed on advice, experience, or referrals.

Related work

Sentiment analysis is an active research area due to its application in opinion mining, emotional detection, and product reviews from online platforms, like Facebook and Twitter.⁹⁻¹¹ It deals fundamentally with the polarity of sentiments of people, events, or topics. Many experiments have been conducted in recent past on health online forums to identify the emotions conveyed in user messages.¹²⁻¹⁴ One of the pioneers in this area is Qiu et al.⁶ They detected community members' sentiment change patterns and found out that most participants change their sentiment positively through interaction. Lexical, and style features were used to create the classification model. The study by Biyani et al.⁷ followed the above research and detected the community's dynamics using domain-dependent and domain-independent features. An improved sentiment classification task by Ofek et al.⁸ was used abstract features derived from dynamic lexicon to reduce the dimensionality. A dynamic sentiment lexicon specifically for the cancer survivor domain was developed, and probabilities of sentiment scores were measured.

Sentiment Analysis has been extensively studied in various domains using deep learning framework. Two studies applied deep CNN to film reviews and achieved a good precision of 86% and 95%.^{15,16} Chen¹⁷ proposed a method for predicting user ratings from user reviews of products or services using deep belief network. Another research by Souma et al.¹⁸ used RNN with LSTM units to explore financial news sentiments' predictive power. Similarly, a recent study conducted SA using a combination of bag of words and CNN on the microblog dataset and film comment dataset.¹⁹ It reported an accuracy of approximately 90%. Some SA tasks in other languages have been significantly improved compared to the conventional methods. A study on Arabic language Hotel reviews using BiLSTM technique reported an accuracy of 83%.²⁰

Two recent studies conducted on medical sentiment analysis on health community forum using CNN and LSTM are reported in cancer community forum.^{21,22} Khanpour and Caragea²³ used a combination of CNN and LSTM to extract emotional support from a breast cancer discussion group and reported an F1 score of 91.4%. Caragea et al.²⁴ used LSTM, CNN and gated RNN to identify optimism and pessimism on online health communities. Empathetic Message in OHCs were found out using combination of CNN and LSTM called ConvLSTM, which outperformed their individual results.²⁵

The above stated studies motivated to frame the main objective of this research work as to find the efficacy of deep learning algorithms in catching the sentiment dynamics compared to conventional machine learning methods. In this aspect, the study was designed to find out:

1. The sentimental features to be considered for classification of patient narratives.
2. The deep learning methodologies to be considered for proper classification of the sentiments.
3. The changes in the sentiment of a patient before and after receiving support.

Methodology

Data source

The dataset used in this study was taken from the Breastcancer.org with the permission of the site administrators. The medical history of users is, by default, private, according to BCO's policy. They may choose to publish them in whole or in part by modifying their privacy settings. Once a user decides to publish, it will become part of the signature to every post they make and visible to anyone reading the post. Nearly, 150,000 posts over the forums of chemotherapy, radiation, and surgery were collected for this study.

Manual annotation

Manual annotation was done by two members of the team trained on annotation strategies. They annotated the same samples of text in parallel. First, read each post, understood its meaning and the emotions expressed in the post, and then they created the label. The mood of the emotions was detected by finding the positive or negative words used in the posts. For instance, the words like "frustrating," "punishing," "disappointment," "sad," "fear" etc. are negative words and "easy," "thank," "grateful," "happy," "lucky" etc. are positive words; the annotators were trained with these kinds of words from positive emotion keywords and negative emotion keywords. Then, their annotations were compared, and a kappa score (agreement score) was computed. A kappa score of 0.63 was achieved. The labels were later reviewed by the clinical expert in the team. In this way, 834 posts were manually labeled and out of which 431 were positive and 403 were negative. A few sample posts and its given labels are shown in Table 1. For instance, due to the presence of words such as "seriously," "frustrating," the first post in the table was given the label "negative" and expressed the negative mood in the post. Similarly, the words "easy" and "without issues" were identified in the second post as "positive" and expressed an overall positive mood in the post.

Overall approach

As stated in section 3.1, nearly 150,000 user narratives (posts) on Breastcancer.org online platform, specifically on chemotherapy, radiation and surgery were considered for the current study, since they are the three most important therapeutic options for cancer treatment and the most nervous or

Table 1. Few sample posts and its label.

	Post	Label
1	had my blood counts done today I l days out from my first round of chemo and just heard back from my onc's nurse. my wbc's are extremely low i am being put on an antibiotic and told no work this week. getting out for fresh air can be tough! this is seriously frustrating!!	Negative
2	my port experience had been easy. i'm run ride swim doing yoga boxing and weight lifting with it without any issues at all! it's took me about a week to get use to it. now i don't even notice that it's there	Positive
3	i am just frustrated. . .anxious because of starting chemo and i am just sick of this crap already (i know i am only at the beginning too!) just feeling spent emotionally the last few days	Negative
4	hey there-i exercised every single day during chemo (even if sometimes all i could manage was a 30 minute walk) and it was fine having a port. i barely noticed. the exercise kept me sane- good luck!	Positive
5	i have a similar issue. i made it thru ac and 4 of 12 taxols. my alts have been high so we skipped the last 4. now my oncologist is saying i am done though i didnt complete all of my treatments. i feel very uncomfortable about stopping early and worry about reoccurrence.	Negative
6	your hair will come back; first sparsely and then more thickly. i gave up my wig 10 weeks after my final chemo and a year after that i have heaps and heaps of hair. Thank god! so be reassured and take care.	Positive

concerned part. Emotional dynamics are highest during the treatment phase and hence the patient's emotions were analyzed on advice, experience, or referrals. Deep learning models were experimented for analyzing sentiment dynamics.

Sentiment analysis is basically a classification problem. This learning methodology requires a large number of annotated data. Since the availability of annotated data was limited, as it was 834 posts, a semi-supervised co-training technique was used to enhance the volume of the training data. With the co-training technique, nearly, 15,000 posts were incrementally labeled and subsequently built a deep learning model. The proposed approach was compared with few existing sentiment analysis techniques conducted on large cancer community online-groups and also analyzed the change in sentiment dynamics before and after receiving support from the community.

Enhancement of the size of the data set

The accuracy and performance of deep learning classifier depends on a large number of training data. However, it is time-consuming and expensive to label large amounts of data manually. semi-supervised learning can be used for automatic labelling of unlabeled data. There are two ways to handle semi-supervised learning problems: self-training and co-training.²⁶ In the self-training approach, the classifier is built with limited labeled instances and predict the label for unlabeled instance.²⁷ In co-training, two separate classifiers should be constructed with two disjoint labeled data subsets. Then these two classifiers are learnt and the most confident prediction on the unlabeled data are used to build the additional labeled training data. However, a study has shown the criteria of disjoint labeled data set can be relaxed by considering two feature sets of data.²⁸ Two separate classifiers trained on two feature sets represent two data views and so the performance is more or less the same. The unlabeled data predicted with the most confident label by both classifiers are moved to labeled data set for the next iteration. That is, the instances predicted with same

label from both the classifiers are moved to labeled data set. The conflicting instances were discarded. For instance, the following post received different labels from both the classifiers:

“ my routine will be 6 cycles one every 3 weeks. my chemo is neoadjuvant meaning i am having it before surgery. in a way i’m scared and in a way i am glad to be getting treatment started. i am prayerful that the chemo will knock this you-know-what out of my body”-The patient is scared about chemo, but at the same time glad about the treatment started.

The above procedure repeated until all the unlabeled data were used up and each classifier benefited from improved training data at each iteration and finally, predicted more confident scores. In the proposed study, co-training technique was used for enhancing the size of data set. Word embedding features and sentiment embedding features were used as two views of co-training data in the present study. A domain-specific lexicon was constructed to capture word embedding features and sentiment embedding features. Following sub-sections shows the creation of domain-specific lexicon and generation of features.

Creation of domain-specific lexicon. A domain-specific lexicon was constructed from manually labeled posts, and then polarity and intensity scores were calculated similar to Ofek’ study.⁸ The labeled posts were tokenized into words, and removed the stop-words. Then the unigram and bigram words, named as terms, that appeared in at least three posts were chosen to form the lexicon. Two frequency scores were calculated for each term, PosFrq (positive frequency), and NegFrq (negative frequency) for each class they belong according to following formula:

$$\text{PosFrq} = \text{No. of positive posts the term appeared} / \text{Total no. of positive posts.} \quad (1)$$

$$\text{NegFrq} = \text{No. of negative posts the term appeared} / \text{Total no. of negative posts.} \quad (2)$$

For instance, if a term “diagnosis” occurs in 20 negative posts, out of 103, and 5 positives out of 118, its NegFrq is 0.195 (20/103), and PosFrq is 0.042 (5/118). Subsequently, the probability score for each class was determined as:

$$\text{PosProbScore (positive probability score)} = \text{PosFrq} / \text{NegFrq} \quad (3)$$

$$\text{NegProbScore (negative probability score)} = \text{NegFrq} / \text{PosFrq.} \quad (4)$$

Therefore, “diagnosis” has PosProbScore as 0.215 (0.042/0.195) and NegProbScore as 4.643 (0.195/0.042). Its intensity score was then calculated as:

$$\text{Intensity score} = \text{PosProbScore} - \text{NegProbScore} \quad (5)$$

Thus “diagnose” was obtained intensity score as -4.428, and polarity as negative. In this way, the intensity and polarity were measured and stored for all terms in the lexicon. Table 2 shows a fragment of domain specific lexicon.

Generation of features from domain-specific lexicon. The first feature vector was built from the lexicon using a word embedding technique, word2vec and the embedding dimension was 100.²⁹ Word

Table 2. An example of 10 terms in lexicon.

Term	Polarity	Intensity
Terrified	Negative	-4.82
Diagnose	Negative	-4.42
Serious infection	Negative	-4.56
Extremely fatigued	Negative	-2.16
Pretty easy	Positive	1.34
Can handle	Positive	2.78
Much better	Positive	3.12
Recommend	Positive	3.62

embedding defines the word on the basis of the terms surrounding the word, in the real number vector. Typically, the embedding dimension in word2vec is between 100 and 300. If it is higher, the training time is slow and the performance at 100 and 200 was more or less the same. Hence, the dimension was chosen to be 100. The Genism library in Python which provides the Word2Vec class was used to operate on a Word2Vec model.

The second type of feature vector was created by the sentiment embedding technique. To capture sufficient sentiment information, a semantic sentiment approach was employed.³⁰ First, most similar terms to a target word was found out using Gensim library. Then, Top-n most similar terms were chosen as seeds according to its cosine similarity value. Rank the Top-n seeds according to its sentiment intensity score contained in the lexicon. Top-k sentimentally similar seeds were selected. A rank is given as 1 for top-ranked seed, 2 for second ranked seed and so on. Table 3 shows, Top-10 similar seeds from Gensim for a target word “terrified” (highlighted in bold in table) ranked according to its sentiment intensity score. As shown, Top-8 seeds were sentimentally similar. The seeds ranked 9th and 10th (highlighted in bold in Table 3) were sentimentally dissimilar and were eliminated. Thus, seeds which were semantically and sentimentally similar were selected.

Finally, a sentiment embedding vector for target word was calculated by the weighted sum of word vectors of those top-k most similar sentiment words according to the following formula:

$$senti_wordvec_i = \frac{1}{k} \sum_{j=1}^k w_j * seed_j \quad (6)$$

where $senti_wordvec_i$ is sentiment vector for target word i , $seed_j$ is word vector for sentimentally similar seed words, w_j is weight of $seed_j$.

Weight of seed words were calculated by the rank of each similar seed words multiplied by its cosine similarity value, so that higher ranked seeds get more weight.

$$w_j = cos_sim_{i,j} * \frac{1}{rank_j} \quad (7)$$

where $rank_j$ is rank of seed word that is, $cos_sim_{i,j}$ is cosine similarity value between target word and each of seed words. In this way, a sentiment embedding vector for each of the terms in the lexicon were constructed.

With these two feature vectors, two classifiers were built for co-training with 80% of positive and negative labeled posts. The remaining 20% of the labeled posts were used for testing. Different machine learning algorithms were experimented for these two feature vectors, such as Naïve

Table 3. Words ranked according to sentiment intensity score.

Words ranked according to cosine similarity	Sentiment intensity score	Rank
Terrified	-4.82	
Frustrated	-5.23	1
Tough	-5.12	2
Seriously frustrating	-4.23	3
Terribly noticeable	-4.11	4
Trouble	-3.54	5
Issue	-3.18	6
Uncomfortable	-2.54	7
Annoying	-2.32	8
Less depressed	1.18	9
Feel better	2.43	10

Bayes, SVM, Logistic Regression, and AdaBoost. To extend the training set, a random sample of 15,000 posts from the unlabeled set was selected. Then a sample of 100 posts from 15,000 posts was given at each iteration to both classifiers for prediction. Posts with the highest confidence label were added to the existing training data. The data set thus enhanced was used for the next training session. The size of the data set was enhanced through such several similar iterations. The performance evaluation of the four classifiers are shown in Performance evaluation section.

Deep learning-based sentiment classifiers

Enhanced with more training data, the efficacy of deep learning classifiers in sentiment analysis was investigated. CNN, LSTM, and BiLSTM's performance were compared using word embedding and sentiment embedding vectors as input representation. All of these classifiers were fed with word vectors. The embedding vector of each word was vector from word embedding technique and sentiment embedding technique, mentioned in the above section. Both the vectors were fed to classifiers separately and compared its impact on final classification's performance. It is shown in "Performance Evaluation" section.

CNN is a deep learning algorithm, similar to a neural network consisting of an input layer, an output layer, and many hidden layers. The architecture of CNN is shown in Figure 1. The input to the CNN is in the form of xnk sentence matrix, where x is the number of sentences, n is the number of words, and k is the dimension of each term's embedding vector.³¹ Thus, each row of the matrix is the embedding vector of a word in a post. The convolutional layer in CNN is the feature extraction layer that uses m filters to every possible window of words in the sentence vector to build a feature map. Feature map is then processed by maxpooling layer to produce summarized features. Finally, SoftMax function is used to classify the posts.

CNN model parameters were varied with the number of convolution layers, the number of filters, and filter size. The experiments were conducted with two and three convolutional layers, 5, 10, 15, 20 filter numbers, and 3 and 4 filter sizes. By evaluating studies in Rani and Kumar,¹⁶ the values of these parameters were defined. The dropout value was 0.5, the number of epochs was 8, and the batch size was 64. With two convolution layers and five filters with a size of 3 yielded a more promising result.

LSTM is a deep learning algorithm, a recurrent neural network (RNN) variant, explicitly designed to learn textual data sequences. Unlike CNN, which ignores word order, LSTM naturally

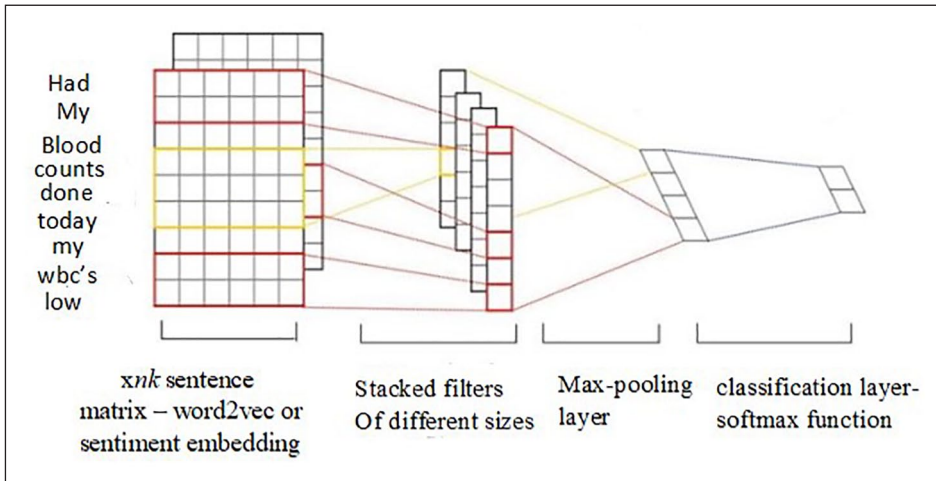


Figure 1. CNN architecture.³¹

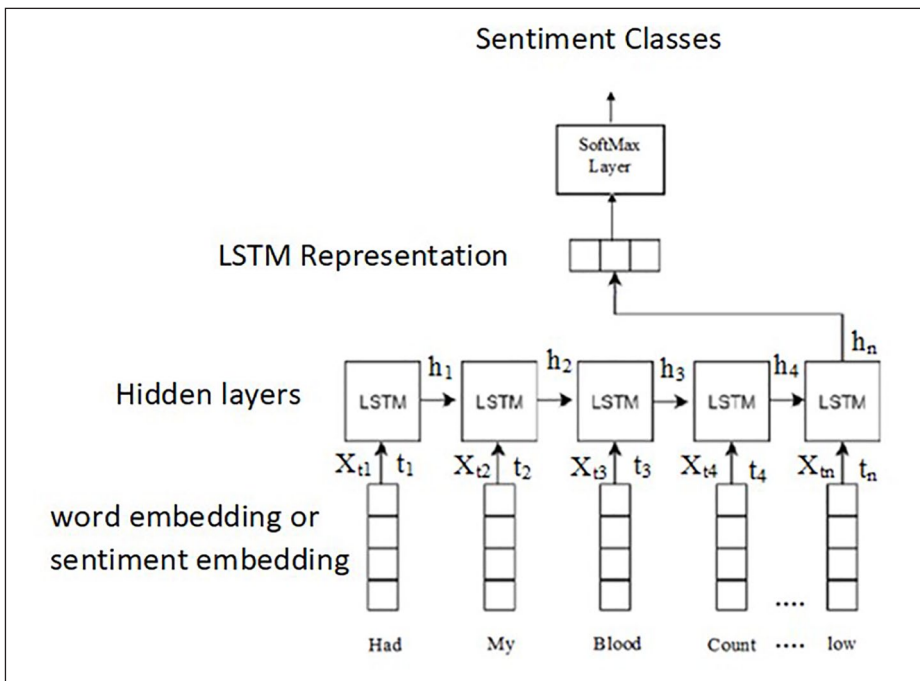


Figure 2. LSTM architecture.³²

manages word order and word sequence. LSTM comprises many hidden layers of gated cells known as memory cells, that either remember or forget information.³² The architecture is shown in Figure 2. The input is a post that contains a series of words, and each word is an embedding vector. At each time step t , the cell value h_t is computed by a linear combination of current input and the previous state as follows:

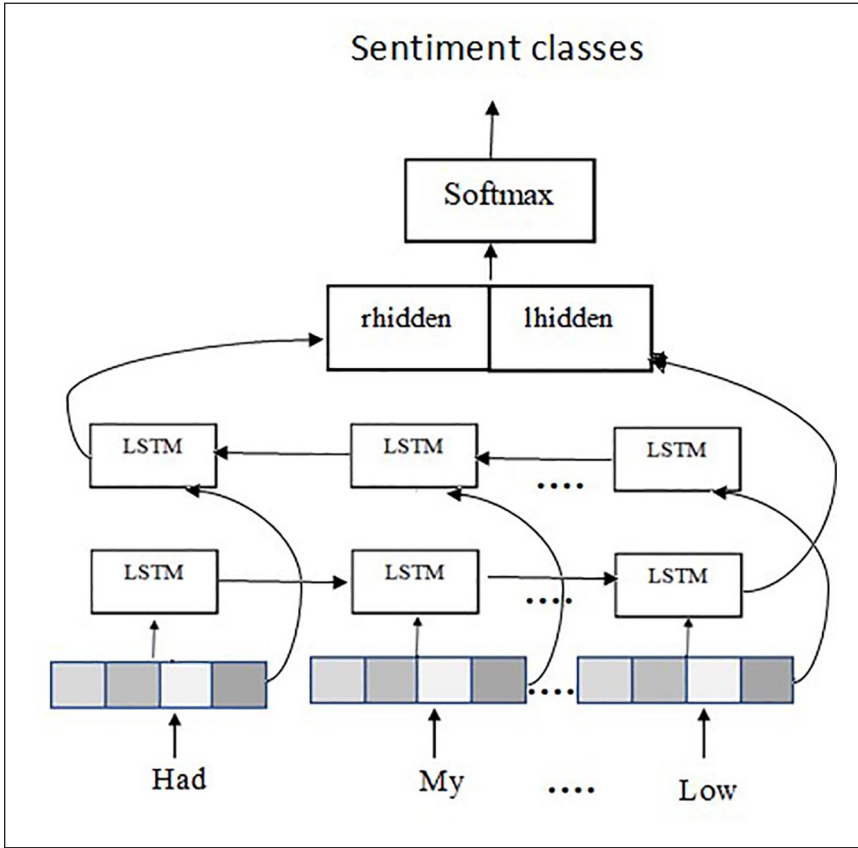


Figure 3. BiLSTM architecture.³³

$$h_t = f(W_h h_{t-1} + W_x X_t) \quad (8)$$

where f is tanh activation function, W_h is the weight of the hidden layer, h_{t-1} is the previous state, W_x is the weight of current input, and X_t is the current input word vector. The maximum length of the sequence determines the number of hidden states. The above equation is repeated for all the hidden states. The SoftMax function is finally applied to the last hidden layer output to obtain the highest probability for positive or negative class.

The hyperparameter settings for LSTM were: post with a minimum length of 25 words and a maximum length of 500 words, the batch size was 128, the number of LSTM units was 256, and the number of epochs was 10 with a learning rate of 0.01 and drop out of 0.5, which yielded promising result.

Since LSTM is a one-directional approach, it cannot get the information from the future sequences of words. BiLSTM maintains two separate forward and backward input states formed by two LSTMs.³³ The architecture of BiLSTM is shown in Figure 3. For each input sequence, the hidden vectors are computed based on past and future data. Past data is extracted from the first LSTM, which is provided with words from left to right, while future data is extracted using words in the opposite direction in the second LSTM. The output value for each word is determined using past feature vector, $lhhidden$ from left LSTM units, and future feature vectors, $rhhidden$ from right LSTM units. Thus, the information surrounding the current input word is effectively captured in

Table 4. Forums and threads.

Forums	Total no of threads	Forum-specific keywords	No of posts having the keywords
Chemotherapy	5899	Port placement, neutropenia, neuropathy, neoadjuvant, missing chemo, cold cap, Adriamycin, mouth sore, hot flashes, nausea, Taxotere	58,471
Surgery	5432	Nerve block, onco test, mastectomy or lumpectomy, numbness, bi-lateral mastectomy, follow-up, reconstruction	50,498
Radiation	3345	Missing radiation, nausea, rashes, reconstruction, pbi therapy, fatigue, rads position, frozen shoulder, physiotherapy, lymphedema, rads after mastectomy, plastic surgeon, nipple surgery, tissue expander	41,769

the BiLSTM. The output vector is then processed by SoftMax layer to obtain the highest probability for sentiment classes. Layers 1, 2, and 3 stacked with 256 units each were experimented. The two-layer BiLSTM packed with 256 units provided the best output.

Sentiment change analysis

Sentiment change analysis was carried out on the labeled posts, to measure the sentiment dynamics and assess satisfaction level. Our analysis was focused on the threads of three forums surgery, chemotherapy, and radiation. Each thread in the forums was opened with the support seeker's post, and was followed by the support provider's answers and self-replies. To monitor the threads with a larger number of discussions, the most commonly spoken terms were extracted, believing these were the most discussed issues and the most popular threads. Then the threads with these terms were collected, and it was interesting to note that support seekers are beginning their discussions with these concerns. The total number of threads, and top threads having the most commonly spoken keywords in each forum and the number of posts having the keywords are shown in Table 4.

The sentiment changes of support seekers with regard to four factors, such as number of replies from providers, number of self-replies, average sentiment of provider's reply and self-replies, and time elapsed between first post seeking support and first reply were examined.⁶ The sentiment change can be measured by the difference between average sentiment of self-replies and initial post.

For any post P , the model predicts the probability of the sentiment class it belongs to as $Pr(P)$. Let sentiment probability of initial post be $Pr(P_i)$, sentiment probability of later self-replies be $Pr(P_{sr})$ and the sentiment probability of providers replies be $Pr(P_{pp})$. Sentiment change, ΔPr , of support seekers can be computed as the difference between average sentiment probability of self-replies and sentiment probability of initial post⁶:

$$\Delta Pr = \frac{\sum_{srj=0}^n Pr(P_{srj})}{n} - Pr(P_i) \quad (9)$$

where $Pr(P_{srj})$ is sentiment probability of self-replies and n is total number of self-replies.

The impact of influencing factors on sentiment change is shown in Performance Evaluation section. The prediction probability of the best deep learning model was used to compute the sentiment change ΔPr .

Table 5. Performance comparison of classifiers used in co-training technique.

Classifiers	Features	Precision	Recall	F-1 score
Naïve Bayes	WE	0.722	0.732	0.745
	SE	0.754	0.723	0.743
Logistic regression	WE	0.710	0.711	0.701
	SE	0.732	0.721	0.721
AdaBoost	WE	0.762	0.764	0.778
	SE	0.782	0.784	0.786
SVM	WE	0.843	0.846	0.845
	SE	0.867	0.859	0.869

Table 6. Performance comparison with previous studies.

Classifiers	Precision	Recall	F-1 score
SVM	0.867	0.859	0.869
Biyani et al. ⁸	0.844	0.843	0.844
Ofek et al. ⁹	0.819	0.823	0.811

Performance evaluation

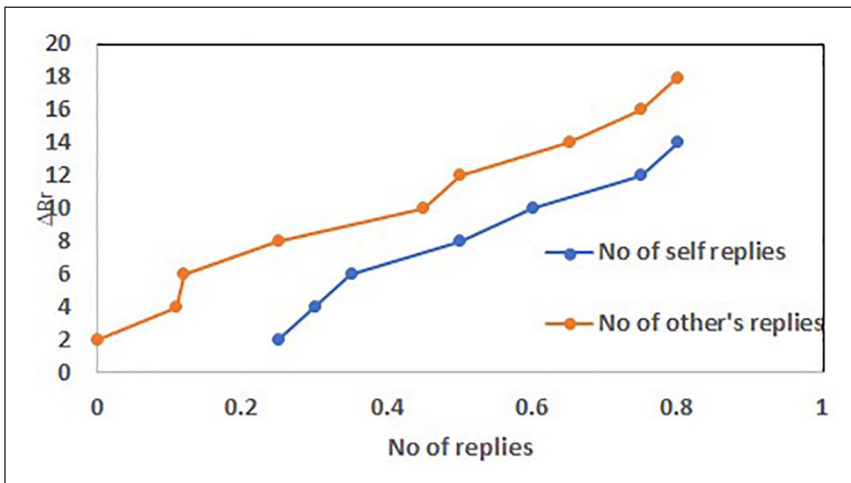
The first experiment was to evaluate the performance of the classifiers used for co-training technique with two feature vectors - word embedding (WE) and sentiment embedding (SE). Out of the 15,000 unlabeled posts, 8432 were labelled as positive and 5110 as negative and the remaining conflicting. A total of 13,542 posts were labelled, of which positive, and 5110 as negative, and the remaining were conflicting. Python's Scikit library was used to implement the classifiers for co-training. Macro-averaged precision, recall, and F1-score were used to evaluate the results. The results are shown in Table 5. The SVM classifier trained on sentiment embedding features gave the most promising F1-score (highlighted in bold in the table) of 86.9%.

Our best classifier with two previous studies was compared. The result is shown in Table 6. The best classifier SVM (highlighted in bold in the table) outperformed the two previous studies. While Biyani et al.⁷ used co-training approach with Bag-of words vector representation, Ofek et al.⁸ used dynamic lexicon with TF-vector representation. As the answer to the SQ1, it was found that SVM classifier with sentiment embedding feature vector from domain-specific lexicon can effectively capture the sentiment features in user's posts.

The second experiment was conducted to compare deep learning classifiers. CNN, LSTM, and BiLSTM with WE and SE as feature vectors were implemented. 80% of the labeled data from the enhanced data set was used for training and 20% for testing with 10-fold cross-validation. All these models were implemented in python using the Keras package. Macro-averaged precision, recall, and F1-score were used for comparison. SVM classifier with SE feature vector, was considered as the baseline model for deep learning algorithms comparison. Table 7 compares the performance of CNN, LSTM, and BiLSTM against the base model SVM classifier. It is evident that BiLSTM with the SE features is a very promising algorithm for classifying the sentiment polarity of user's posts. BiLSTM yielded F1-score of 91.9% which is highlighted in bold in the Table.

Table 7. Performance comparison of deep learning algorithms with base model.

Method	WE (word embedding)			SE (sentiment embedding)		
	Precision	Recall	F1-score	Precision	Recall	F1-score
SVM	0.843	0.846	0.845	0.867	0.859	0.869
CNN	0.858	0.859	0.859	0.872	0.875	0.877
LSTM	0.898	0.897	0.898	0.901	0.902	0.902
Bi-LSTM	0.903	0.905	0.903	0.912	0.916	0.919

**Figure 4.** Sentiment change ΔPr with no of replies.

The third experiment was to find out sentiment dynamics. The proposed BiLSTM model was applied over 150,000 posts belonging to different topic's threads. The initial posts in various threads were from support seekers having negative sentiment polarity as they were more anxious or worried about several reasons. As stated in section 3.4, the number of replies from support providers, the number of self-replies, the average sentiment of these replies, and time elapsed between the first post and first reply were computed. The number of replies from support providers indicated interest in topics, and the number of self-replies indicated self-involvement in topics. Sentiment change was computed using equation (9). Figure 4 shows the sentiment changes in these threads to the number of responses. The sentiments of Self-replies' started to change when the number of others' replies increased. The influence of the average sentiment of supporters' replies on the sentiment change of support seeker's post is shown in Figure 5. It shows that higher the average sentiment score of replies, higher the sentiment change score of the seeker. The Figure 6 shows the relationship between sentiment change and time elapsed. Timely reply only leads to improvement in the sentiment change of support seekers. Another significant point is that a late response from supporters has a zero or negative effect on the sentiment of the seeker.

The sentiment change scores of various concerns in the three forums are plotted in following figures. It is evident from Figure 7(a) that, on the chemotherapy forum, for the concerns "port placement" and "cold cap," the sentiment change was highly positive, while for concerns "missing

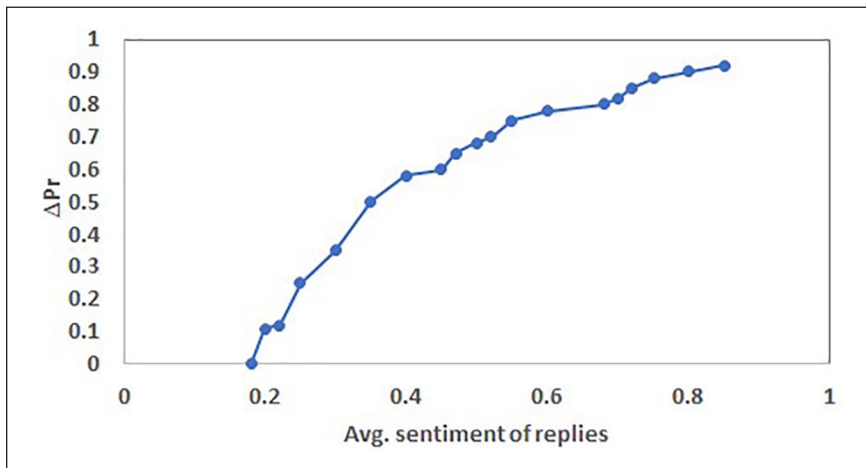


Figure 5. Sentiment change ΔPr with average sentiment of replies.

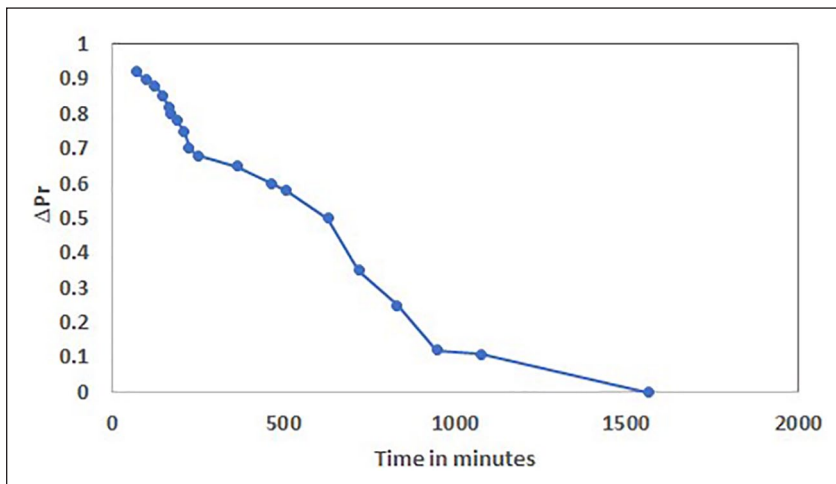


Figure 6. Sentiment change ΔPr with time elapsed between first post and first reply.

chemo,” “hot flashes,” and “nausea,” the sentiment change was negative. From Figure 7(b) it is evident that, on the surgery forum, for the concern “follow-up,” the sentiment change was highly positive, while for the concern “nerve block,” the sentiment change was negative. Figure 7(c) shows, for the concerns “Physiotherapy” and “frozen shoulder” the sentiment change was highly positive while, for the concerns “rads position,” “lymphedema,” and “tissue expander” the sentiment change was negative.

From the sentiment change value of each topics, an assessment of satisfaction level was performed. The level of satisfaction was computed as follows:

If $0.5 < \Delta Pr < 1$, then satisfaction is highly increased, if $0 < \Delta Pr < 0.5$, then satisfaction is slightly increased and if $-0.5 < \Delta Pr < -1$, then satisfaction is highly decreased and if $-0.5 < \Delta Pr < 0$, then satisfaction is slightly decreased. The result is tabulated in Table 8.

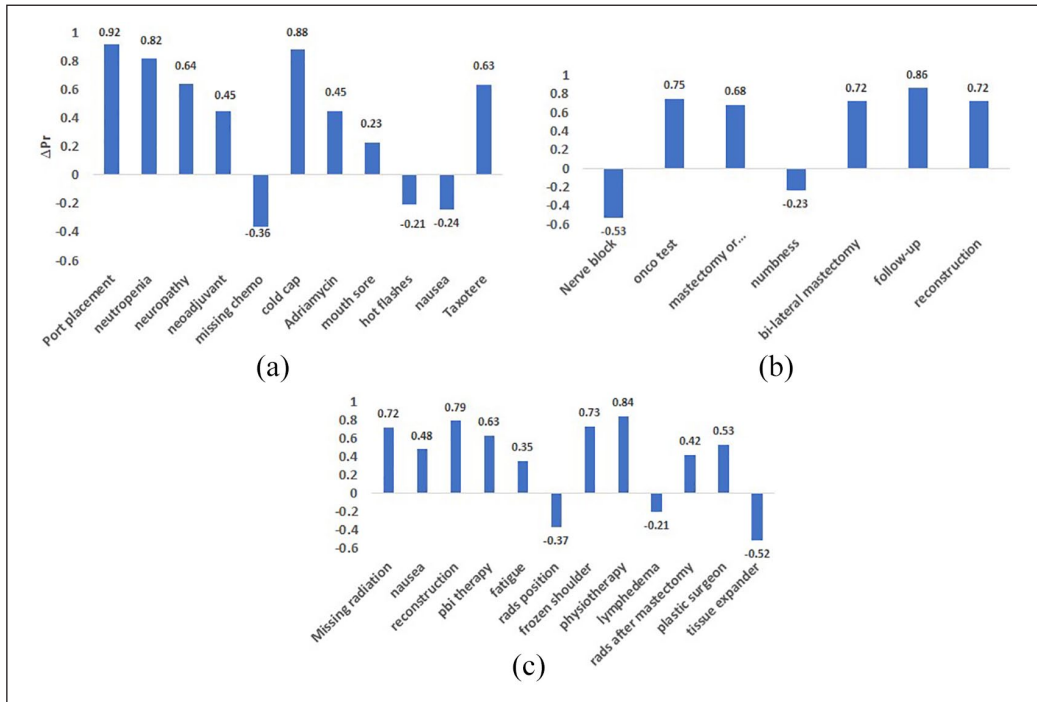


Figure 7. Sentiment changes in forums: (a) sentiment changes on chemotherapy forum, (b) sentiment changes on surgery forum, and (c) sentiment changes on radiation forum.

The studies have shown that the topics with highest satisfaction level obtained great support from other experienced people. Their experience, opinion or advice made positive impact on the seekers. Some side effects of treatments like “mouth sore,” “nausea,” “lymphedema,” “fatigue” obtained slightly increased or decreased satisfaction level. This may be because the side effects were a significant problem for them, even at the advice of the supporters. It can also be seen that satisfaction level had highly decreased with “tissue expander” topic discussed at the radiation forum. It may be because of bad experience of other patients during radiation. The highly or slightly decreased satisfied topics require further assistants from site moderators or health care practitioners.

Discussion

Sentiment analysis (SA) in OHCs helps to catch emotions in opinions, advice, or referrals. Detecting the sentiment features in the text is very vital in efficiently classifying the polarity of sentiment. Unlike traditional machine learning approaches, deep learning algorithms would identify sentiments more effectively. Identifying the change in sentiments of patients during the course of treatment will facilitate insight into the matters that require further focus on long-term survival.

Generally, sentiment analysis uses various sentiment lexicons, such as SentiWordNet, SenticNet, or LIWC linguistic tool, etc. However, the general sentiment lexicon is not ideal for OHCs due to the occurrence of domain-based words. For instance, in the general context, the term “positive” has positive polarity, but this may be a negative polarity for OHCs. Therefore, our domain-dependent

Table 8. Satisfaction level of different topics in Forums.

Satisfaction level			
Highly increased	Slightly increased	Slightly decreased	Highly decreased
Port placement	Neoadjuvant		Tissue expander
Neutropenia	Missing chemo	Hot flashes	
Neuropathy	Adriamycin	Nausea	
Cold cap	Mouth sore	Numbness	
Taxotere		Rads position	
Nerve block	Fatigue	Lymphedema	
Onco test	Rads after mastectomy		
Mastectomy or lumpectomy			
Bi-lateral mastectomy			
Follow-up			
Reconstruction			
Missing radiation			
Nausea			
Reconstruction			
Pbi therapy			
Frozen shoulder			
Physiotherapy			
Plastic surgeon			

lexicon adapted for the breast cancer community is ideally suited for capturing the patient's sentiment. However, an existing study had used a domain specific-lexicon with abstracted features as a feature vector to minimize dimensionality.⁸ The sentiment embedding technique in our analysis, not only embedded word vectors into a low-dimensional vector space but also detected contextual as well as sentimental meaning. For example, the word "terrified," had two terms "feel better" and "less depressed" in closer proximity according to its cosine similarity score. But those terms were eliminated by sentiment embedding technique due to its opposite sentiment polarity. Our results confirm the efficacy of deep learning algorithms to classify the sentiment polarity in the user's narration at the onset. The effectiveness of LSTM and BiLSTM in learning the data sequences of textual data, over CNN is proved once again. In fact, BiLSTM has provided the highest result due to its capability to detect future and past word knowledge, which is not feasible in the LSTM algorithm.

This work leads to a better understanding of the various needs of patients during the treatments. Their various concerns/needs during chemotherapy, radiation, or surgery are evident in the most common keywords shown in Table 4. Sentiment change analysis of the posts containing these concerns indicates their satisfaction level, after getting support from similar patients which is shown in Table 8. Certain issues are highly resolved/getting satisfactory advice. For example, the experience and advice from supporter's toward "port placement," "reconstruction," or "cold cap" are more encouraging. A positive change in text sentiment means an increase in user's satisfaction. This is evident in the following Table 9 that shows a thread snippet asking about lump during chemo. First post is about feeling lump on unaffected breast during chemo and expressing negative emotion. Later her sentiment is changed (in the third post) and became satisfied with the advice from the other fellow patient (second post) and was able to take informed decision.

Table 9. A thread snippet.

i have just finished my third dose of chemo (taxotere and cytoaban). have one more dose to go. since i am undergoing the chemo i assume you cannot develop any cancer in your breasts because the chemo is killing the cancer cells. today i think i felt a lump or two in the breast that didn't have the breast cancer. I'm much worried. is this possible being on chemo? thanks in advance for any info on this

hi paintedlady i just wanted to share my own experience - i felt a lump in my "good" breast after my 4th round of taxotere-carboplatin-herceptin. i absolutely freaked out went to my onc the next day and he sent me for an ultrasound that afternoon. it turned out to be a cyst. i hope that's the case with you - but you should get it checked out asap just in case. chances are it's not cancer but go get an ultrasound to make sure. good luck!

thank you so much for your feedback. i was freaking out when i felt the lump in my unaffected breast. hearing stories like yours give me hope, I'll get it checked asap.

Certain issues still need to be addressed to get a better feeling. For example, “missing chemotherapy schedule” or “neoadjuvant chemo” are some of the questions that still need more attention. “Fatigue,” “nausea,” or other side effects may be the reason for missing chemo schedules. But their concerns over missing schedule need to be addressed by clinical staff. The concerns they raise like “neoadjuvant chemo” leads to the fact that more education is to be given to the patients from healthcare providers. One of the most concerning elements of surgery is taking decisions over “bilateral mastectomy or mastectomy” or “lumpectomy”. They spoke more about “reconstruction” during the radiation process and got satisfying advice. Another conclusion is that “reconstruction,” “tissue expander,” and “plastic surgeon” are talked together during radiation. Very low sentiment changes toward “tissue expander” may be due to low sentiments from the experienced people. It may be because of bad experience of other patients during radiation. These findings are an essential reference for healthcare providers to gain more insight into patient’s real matters and to then provide further assistance.

Our experiment was limited to only three forums, but we intend to expand it to other forums in future to obtain awareness in other areas of concerns. Another limitation of our study was the small size of manually annotated posts. Though the study could enhance labelled training data automatically, there were still some issues with conflicting posts. As seen in the case of the conflicting post in section 3.4, the posts having both emotions need to be more examined. In these types of posts, the prominent sentiment must be accurately identified. Since our domain-specific lexicon was adapted to the breast cancer community, further studies are necessary to determine the lexicon’s effectiveness in other health community domains. As this study discussed the binary classification of sentiment, multi-level sentiment classification is also an important path for future studies.

Conclusion

OHCs are an excellent platform for sharing the patients’ experiences, feelings, and practical tips for coping with daily challenges during treatment. This study investigated the efficacy of deep learning algorithms to detect the sentiment change in breast cancer patients’ narrations. Based on the SVM binary classifier, we compared the effect of word embedding and sentiment embedding for vectorization on sentiment classification. The co-training approach with sentiment embedding is a promising method for increasing the training data set, which is very important for deep learning classifiers’ success. Finally, our BiLSTM model outperformed all the existing studies and baseline model with an F1-score of 91.9%. The ultimate goal was to assess users’ satisfaction level after getting support and find out which matter, they raise has to provide further attention. The results show that there are various concerns still patients need immediate attention from moderators of sites or healthcare providers. These findings can be beneficial in any other health community forums to detect the future needs and subsequent long-term survival of patients.

Acknowledgment

The authors would like to acknowledge Breastcancer.org Community (website) as well as the research students from the BioHealth Informatics Department, IUPUI, who provided support in annotation part. Authors would also like to acknowledge Dr. Mathew J Palakal, Associate Dean, School of Informatics and Computing, IUPUI, Indianapolis, US, for the useful interaction and constructive ideas during the work.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Athira Balakrishnan  <https://orcid.org/0000-0001-7830-2461>

References

1. Barraclough J. *Cancer and emotion: a practical guide to psycho-oncology*. 3rd ed. England: John Wiley & Sons, 1999.
2. Bambina A. *Online social support: the interplay of social networks and computer-mediated communication*. New York: Cambria Press, 2007.
3. Wang X, Zhao K and Street N. Social support and user engagement in online health communities. In: Zheng X, Zeng D, Chen H, et al. (eds) *International Conference on Smart Health*. Cham: Springer, 2014, pp.97–110.
4. Mirošević Š, Prins JB, Selič P, et al. Prevalence and factors associated with unmet needs in post-treatment cancer survivors: a systematic review. *Eur J Cancer Care (Engl)* 2019; 28(3): e13060.
5. Biyani P, Caragea C, Mitra P, et al. Identifying emotional and informational support in online health communities. In: *Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers*, Dublin, Ireland, 23–29 August 2014, pp.827–836. Ireland: Dublin City University and Association for Computational Linguistics
6. Qiu B, Zhao K, Mitra P, et al. Get online support, feel better—sentiment analysis and dynamics in an online cancer survivor community. In: *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*, 9 October 2011, pp.274–281. New York: IEEE.
7. Biyani P, Caragea C, Mitra P, et al. Co-training over domain-independent and domain-dependent features for sentiment analysis of an online cancer support community. In: *2013 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2013)*, 25 August 2013, pp.413–417. New York: IEEE.
8. Ofek N, Caragea C, Rokach L, et al. Improving sentiment analysis in an online cancer survivor community using dynamic sentiment lexicon. In: *2013 international conference on social intelligence and technology*, 8 May 2013, pp.109–113. New York: IEEE.
9. Ly DK, Sugiyama K, Lin Z, et al. Product review summarization from a deeper perspective. In: *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, 13 June 2011, pp.311–314. New York: ACM.
10. Neri F, Aliprandi C, Capecci F, et al. Sentiment analysis on social media. In: *2012 IEEE/ACM international conference on advances in social networks analysis and mining*, Istanbul, Turkey, 26–29 August 2012, vol. 12, pp.919–926. New York: IEEE.
11. Jiang L, Yu M, Zhou M, et al. Target-dependent twitter sentiment classification. In: *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, Portland, Oregon, 19–24 June 2011, pp.151–160. United States: Association for Computational Linguistics.
12. Wiley MT, Jin C, Hristidis V, et al. Pharmaceutical drugs chatter on online social networks. *J Biomed Inform* 2014; 49 :245–254.
13. Zhang L, Hall M and Bastola D. Utilizing Twitter data for analysis of chemotherapy. *Int J Med Inform* 2018; 120: 92–100.

14. Zhang S, Bantum E, Owen J, et al. Does sustained participation in an online health community affect sentiment? *AMIA Annu Symp Proc* 2014; 2014: 1970–1979.
15. Dos Santos C and Gatti M. Deep convolutional neural networks for sentiment analysis of short texts. In: *Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers*, Dublin, Ireland, August 2014, pp.69–78. Ireland: Dublin City University and Association for Computational Linguistics
16. Rani S and Kumar P. Deep learning based sentiment analysis using convolution neural network. *Arab J Sci Eng* 2019; 44(4): 3305–3314.
17. Chen RC. User rating classification via deep belief network learning and sentiment analysis. *IEEE Trans Comput Soc Syst* 2019; 6(3): 535–546.
18. Souma W, Vodenska I and Aoyama H. Enhanced news sentiment analysis using deep learning methods. *J Comput Soc Sci* 2019; 2(1): 33–46.
19. Liu B. Text sentiment analysis based on CBOW model and deep learning in big data environment. *J Ambient Intell Humaniz Comput* 2020; 11(2): 451–458.
20. Al-Smadi M, Talafha B, Al-Ayyoub M, et al. Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *Int J Mach Learn Cybern* 2019; 10(8): 2163–2175.
21. Yadav S, Ekbal A, Saha S, et al. Medical sentiment analysis using social media: towards building a patient assisted system. In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Japan, May 2018.
22. Edara DC, Vanukuri LP, Sistla V, et al. Sentiment analysis and text categorization of cancer medical records with LSTM. *J Ambient Intell Humaniz Comput*. Epub ahead of print 16 July 2019. DOI: 10.1007/s12652-019-01399-8.
23. Khanpour H and Caragea C. Fine-grained emotion detection in health-related online posts. In: *Proceedings of the 2018 conference on empirical methods in natural language processing*, Brussels, Belgium, 31 October–4 November 2018, pp.1160–1166. United States: Association for Computational Linguistics.
24. Caragea C, Dinu LP and Dumitru B. Exploring optimism and pessimism in twitter using deep learning. In: *Proceedings of the 2018 conference on empirical methods in natural language processing*, Brussels, Belgium, 31 October–4 November 2018, pp.652–658. United States: Association for Computational Linguistics.
25. Khanpour H, Caragea C and Biyani P. Identifying empathetic messages in online health communities. In: *Proceedings of the eighth international joint conference on natural language processing*, Taipei, Taiwan, 27 November–1 December 2017, vol. 2, pp.246–251. AFNLP.
26. Keyvanpour MR and Imani MB. Semi-supervised text categorization: exploiting unlabeled data using ensemble learning algorithms. *Intell Data Anal* 2013; 17(3): 367–385.
27. Blum A and Mitchell T. Combining labeled and unlabeled data with co-training. In: *Proceedings of the eleventh annual conference on Computational learning theory*, 24 July 1998, pp.92–100. New York: ACM.
28. Balcan MF, Blum A and Yang K. Co-training and expansion: towards bridging theory and practice. In: Saul LK, Weiss Y and Bottou L (eds) *Advances in neural information processing systems*. Cambridge, MA and London: MIT Press, 2005, pp.89–96.
29. Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. 16 January 2013.
30. Yu LC, Wang J, Lai KR, et al. Refining word embeddings using intensity scores for sentiment analysis. *IEEE/ACM Trans Audio Speech Lang Process* 2017; 26(3): 671–681.
31. Yao C, Qu Y, Jin B, et al. A convolutional neural network model for online medical guidance. *IEEE Access* 2016; 4: 4094–4103.
32. Vo QH, Nguyen HT, Le B, et al. Multi-channel LSTM-CNN model for Vietnamese sentiment analysis. In: 2017 9th international conference on knowledge and systems engineering (KSE), 19 October 2017, pp.24–29. New York: IEEE.
33. Chen CW, Tseng SP, Kuan TW, et al. Outpatient text classification using attention-based bidirectional LSTM for robot-assisted servicing in hospital. *Information* 2020; 11(2): 106.