An answer recommendation framework for an online cancer community forum

B. Athira[1] · Sumam Mary Idicula[2] · Josette Jones[3] · Anand Kulanthaivel[3]

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Health community forums are a kind of online platform to discuss various matters related to management of illness. People are increasingly searching for answers online, particularly when they are diagnosed with cancer like life-threatening diseases. People seek suggestions or advice through these platforms to make decisions during their treatments. However, locating the correct information or similar people is often a great challenge for them. In this scenario, this paper proposes an answer recommendation system in an online breast cancer community forum that provide guidance and valuable references to users while making decisions. The answer is the summary of already discussed topic in the forum, so that they do not need to go through all the answer posts which spans over multiple pages or initiate a thread once again. There are three phases for the answer recommendation system, including query similarity model to retrieve the past similar query, query-answer pair generation and answer recommendation. Query similarity model is employed by a Siamese network with Bi-LSTM architecture which could achieve an F1-score of 85.5%. Also, the paper shows the efficacy of transfer learning technique to generalize the model well in our breast cancer query-query pair data set. The query-answer pairs are generated by an extractive summarization technique that is based on an optimization algorithm. The effectiveness of the generated summary is evaluated based on a manually generated summary, and the result shows a ROUGE-1 score of 49%.

Keywords Answer recommendation · Breast cancer community · Query similarity · Summarization

✉ B. Athira
   athiramalu@gmail.com

   Sumam Mary Idicula
   sumamdavid123@gmail.com

   Josette Jones
   jofjones@iupui.edu

   Anand Kulanthaivel
   akulanth@iu.edu

[1] Deparment of CS & IT, School of Computing, Amrita Vishwa Vidyapeetham, Kochi, India

[2] Department of Computer Science, Muthoot Institute of Technology and Science, Kochi, India

[3] BioHealth Informatics Department, IUPUI, Indianapolis, IN 46202, USA

1 Introduction

Health care is gradually shifting from clinician to patient-centered. The patient's role has changed from passive health information recipients to active information seekers and even information providers. Patient decision-making is a critical component of patient-centered health care. The World Health Organization has stated that patient involvement in care is not only desirable but also a social, economic, and technical necessity [42]. A patient's most important type of support is to obtain suggestions and information when they make decisions during their treatment [23]. However, clinicians tend to focus on the clinical impact of disease and may ignore patients' emotional well-being and daily life.

In today's digital world, it is increasingly common for patients to join online health communities (OHCs) to share information and acquire knowledge and support each other users. OHCs contain multiple message boards or forums used by group of users to share concerns about health problems and needs [14]. The goal of such communities is to support patients with chronic conditions and provide them with the opportunity to understand their medical conditions [28]. The user's aim is to interact with patients who have common problems to share their health conditions and know how they overcome similar situations. OHCs contain enormous amounts of patient-generated content in the form of threads on various health-related topics. A thread is started by a person with a question, and responses are posted to that thread. Many patients find help in decision-making by asking questions as their illness progresses. If the patient can locate the relevant information that is already discussed, they can experience a quick and effective information search, increase their level of satisfaction, and thereby make a quick and informed decision. Nevertheless, an information search is a significant challenge for patients attempting to locate relevant information from experienced persons within the OHCs owing to information overload. Therefore, an answer recommendation can overcome this challenge.

In general, a recommendation system recommends items or products to users based on their interests or preferences. A conventional recommendation system relies on two techniques: content-based technique and collaborative filtering technique [2]. Whereas content-based techniques recommend items based on the description of items and user preferences, a collaborative-based technique recommends the items based on similar user preferences. For example, in traditional question-answering forums like Stack Overflow or Quora, the best answers are recommended to a user query based on the voting of community members [39]. Rating/voting is an essential feature used by commercial websites/forums to recommend items [45]. However, traditional recommendation techniques cannot be applied to OHCs owing to the absence of rating/voting information and the lack of a "best answer," and only experiences and opinions are shared instead. On these grounds, an automatic answer recommendation system in the current study provides valuable information and suggestions from posts of experienced patients responding to individual queries.

For the OHC, the current study considered a large breast cancer forum, Breastcancer. org, for recommending answers to patient queries. Whereas early detection and timely and efficient care for breast cancer are increasingly made available, it is also essential to learn how to handle the condition and preserve the quality of daily life. The factors associated with unmet needs in post-treatment cancer survivors were identified in [26]. The largest factor identified is 'being informed about the things one can do to help yourself to get well'. Hence, seeking and providing support for similar patients is a critical requirement for patients living with cancer. A recent work [3] presents a machine learning approach for discovering health care services created by multiple stakeholders in a social media group. They examined Twitter data on the Arabic language in cancer diseases. Another study [7]

looked at the sentiment dynamics of cancer patients in a social media forum by analyzing the patients' posts. A large number of patients have joined any of these forums and have discussed various concerns. However, a user having a similar problem to that discussed in a thread might not want to read all of the posts. Often, threads contain several answer posts and the majority of text includes sympathy as well as stories. Information search is a significant challenge for them when they attempt to locate relevant information within these contents. Hence, user may prefer a brief summary of discussion. Additionally, they prefer to get relevant information according to their specific requirements. So, summarizing these posts is a necessity.

The current study consists of three phases to recommend answer to patient queries. The first phase of answer recommendation is to find a similar query in the threads. The second phase is to summarize the answer posts and third phase is to recommend the answer. A deep learning architecture, a Siamese network, is used to find the closest query. If similar query is not found, the patient can initiate a new thread/or the moderator of the forums is requested to respond. Summarization techniques used for an answer recommendation include an extractive summarization.

The main objective of the present study is to recommend an optimal answer to a user query, if similar concern has already been raised and is discussed in the forum. To recommend the answer, the following tasks are executed.

i)   Build an effective model to find similar query from the repository by effectively capturing the medical knowledge contained in the query.
ii) Generate a summary from the answer posts in the thread by optimally selecting the most informative sentences without losing the relevant information.

The rest of the paper is organized as follows: Section 2 describes previous studies on summarization and query similarity tasks. Section 3 explains the overall architecture of the system, query similarity model and summary generation technique. Section 4 discusses the experiment and results analysis. Finally, in Section 5, the discussion of the study are described, followed by the conclusion and future direction.

2 Related studies

Generally, the application of recommender system in healthcare targets two types of end users, i.e., patients and healthcare professionals [43]. Whereas, health professionals' benefit from clinical guidelines or research articles in treatment and diagnosis, patients receive suggestions such as diet, exercise, life style recommendations. For instance, using personalized content, Farrell et al. [15] recommended life style changes and Roitman et al. [33] recommended patient safety. The authors of [13, 34] targeted diabetic patients to improve their eating, exercise, and sleeping habits. The authors of [36] highlighted the importance of Health Recommender System (HRS) and proposed a system for assisting in the decision - making processes for both patients and physicians.

To the best of our knowledge, only few studies related to the answer recommendation in healthcare field have been reported. In [41] the authors discussed a medical community question answer system in which the best answer post from a collection of answer posts to a similar question is recommended by considering the quality of the answers, instead of summarizing the answer posts. However, the authors in [9] recommended best posts as summary

using extractive summarization technique based on textual features of posts. By contrast, the current study generates a summary by selecting relevant sentences from all the answer posts.

In general, the techniques for text summarization are extractive or abstractive. In an extractive summarization, key sentences are selected from the original document while keeping the sentence intact, whereas with abstractive summarization new sentences are constructed from the original text by understanding the content [6, 8]. Hence, abstractive summarization is complex. Summarization can also be categorized as either generic or query dependent [32]. Generic summarization contains the major content of document, whereas query-based summarization generates content most relevant to a given query. Because discussion forums started with queries from the user, our approach to summarization in the current study was query-based and the technique used was an extractive summarization.

It is important to determine the semantic relatedness between the current query and archived queries, since the recommendation is the most relevant answer to a user query. Several techniques have been used in the literature to measure the textual semantic similarity. Although earlier researchers have relied on linguistic features [16], focus has shifted to neural network techniques. The introduction of word embedding model is a revolutionary approach to a semantic similarity task, including, cosine similarity in sent2vec [30], Word2vec, Doc2vec [21], Glove vectors [4]. The latest addition to these types of embeddings is contextual word embeddings such as BERT and ELMO. Devlin et al. [12] reported that contextual embedding is more effective than the above word embeddings in textual entailment task. It is because, whereas word2vec or doc2vec generate the same vector for the same word in different contexts, BERT generates distinct vectors for the same word in different contexts. BERT's key benefit is that it uses bi-directional learning to get word context, which means it understands word context by reading it both ways from left to right and right to left at the same time. Following recent trends, several deep learning architectures such as GRU, RNN [20], Tree-LSTM [38] have been employed to solve the semantic similarity task. Yet, another study in [27] used two identical LSTM networks, called Siamese network, to measure the semantic relatedness between sentences. It was reported that more promising results are yielded, than in other neural networks. However, these studies were based on large amount of training data, that is, large number of semantically related textual data. For Instance, the studies in [10, 11] are such types of question-question matching task on online user forums. However, these approaches do not perform well in online health community forums owing to the presence of many medically related terms. In addition, there are no large medical query-query pair datasets available. This challenge was recently handled in [25] by applying a transfer learning technique, in which a double fine-tuning approach was implemented. First, the model was trained with large medical question-answer data set and then, fine-tuned with a small question-question data set. Hence, to capture the medical knowledge contained in the query, the current study adopted the concept of transfer learning with a double fine-tuning approach.

To generate the most concise summary related to query, several optimization techniques have been considered in the literature. Relevant sentence extraction, redundancy between sentences and content coverage are the critical issues in forming the summary. Genetic algorithm (GA) [17, 46], differential evolution (DE) [19], and particle swam optimization [5, 31] are some of the optimization techniques proposed. Rautray and Balabantaray [32] proposed a Cuckoo search algorithm for multi-document summarization. However, these are all metaheuristic approaches that require large number of parameter tuning and higher computational effort. This motivates us to develop an optimal sentence selection method that contains minimal parameter tuning. In [40], the authors proposed an optimal combination of sentence scoring methods to rank the sentence to be included in the final summary.

Based on the findings of all of the above studies, we determined that providing the best possible response to a user question is critical. To do this task, we must first develop an effective model for locating a query that is similar in the repository. It's also crucial to capture the medical knowledge provided in the query. In the repository, we need to store an optimal answer for each query. To fulfil this objective, the answer must be derived from the reply postings in the respective thread. As a result, it is necessary to choose the most informative sentences without losing essential information.

3 Methodology

This section describes the system for recommending answer to a particular query related to breast cancer. The recommended answer is the summary of a discussion in the forum by experienced patients, moderators of forums and survivors. These answers contain experiences, suggestions, solutions, viewpoints and above all, how they managed a particular situation. The Siamese network with transfer learning technique is implemented to find the similar query. For each similar query, an answer is generated, using extractive summarization technique with an optimization algorithm, and stored in the repository.

3.1 Dataset

The dataset used for the study was taken from a large OHC, Breastcancer.org [18], a platform mainly formed by breast cancer patients, survivors and caregivers. The site is organized into different forums, each consisting of hundreds of threads. The forums considered in the study dealt with "stage I," "stage II, "stage III," "stage IV," "chemotherapy," "radiation," "surgery," "reconstruction," "DCIS," "ILC," "triple negative," "HER2+," and "employment and insurance" etc. Although some of the forums had fewer threads, we took into account all of the threads from these forums. Each thread starts with a user query and the answers made by survivors and caregivers in the form of posts. Each thread contains hundreds of posts. However, there were some queries/threads, that were replied with one answer. Some were considered according to their significance. On average, a query consists of 5 answer posts. There was an average of five sentences per post. The medical history of users is, by default, private according to the site's policy. Permission to collect these posts was granted by the site administrators and posts were scrapped with python and stored in the database. Nearly, 500 queries and their answers were collected from the threads. Two graduate students in health informatics with clinical backgrounds, trained on summarization strategies, manually created the summary. They read each of the answer posts in the thread and extract the relevant sentences as summary. The sentences that differ in meaning and are more similar to the query were chosen. The summary was later validated by clinical experts in the team. The manually created summary is here referred to as, Breast Cancer-Query-Answer pair (BC-QA) dataset. A sample query, its few answers and manually created summary are shown in Table 1.

For the query similarity check, the data set was the same 500 queries that were used for summarization task. For each query, a similar and a dissimilar query pair was created under the following condition: If the thread, from where the query was taken, consists of similar query, then that was considered as similar one. It is because mostly in the same thread during the course of discussion, users may ask similar questions in different way. If it doesn't contain any such query, rewrite the original query in different way by changing the structure of sentence as much as possible without losing

Table 1 A sample query and its manual summary

Query: posting for a friend. a dear friend who is dealing with Breast cancer [stge2] has been left with mega neuropathy in his feet. does anyone have any experience with any drugs that help with the neuropathy or even any words of encouragement?

Answers:

hi: i have continuing neuropathy after taxotere in 2008 - i know some people are helped by taking gabapentin (you'll have to find out the drug name online) . i took one pill it knocked me out for 20 hours - sorry but you can't feel your feet if you're unconscious - my neuropathy has gotten a little better but won't progress any further i am told. your friend should know that exercising makes it better so does soaking your feet in warm water with epsom salts. temporary to be sure but it's all we've got right now. the exercise is important though because it not only makes your feet feel better it makes you much less depressed. perhaps someone else will chime in soon with a better idea. sandy

the gabapentin is Neurontin the later med pregabalinor lyrica is much much better for my neuropathy

i have neuropathy in my feet pre-cancer that's from an unknown cause (idiopathic). the chemo worsened it so i'm exercising to keep it from going any further up my leg. it's important to keep the muscles strong in the legs and feet for balance. drugs for numbness won't bring back feeling in my understanding but they can help with annoying tingling sensations or pain. that's what my neurologist told me. i'm looking into a vitamin based (but needs an rx) supplement called metanx not sure if it will work any better than just taking b6 and amino acid supplements. hopefully your friend's neuropathy which is chemo-induced will lessen and maybe disappear. please tell him to check the bottom of his feet every day to make sure he hasn't injured them or stepped on anything sharp and always wear a soled shoe/slipper. my husband is a diabetic and has to do this too. numb feet need extra attention. best wishes to your friend!

Summary: i know some people are helped by taking gabapentin (you'll have to find out the drug name online). your friend should know that exercising makes it better so does soaking your feet in warm water with epsom salts. the gabapentin is Neurontin the later med pregabalinor lyrica is much much better for my neuropathy. it's important to keep the muscles strong in the legs and feet for balance. drugs for numbness won't bring back feeling in my understanding but they can help with annoying tingling sensations or pain. hopefully your friend's neuropathy which is chemo-induced will lessen and maybe disappear. please tell him to check the bottom of his feet every day to make sure he hasn't injured them or stepped on anything sharp and always wear a soled shoe/slipper.

the meaning. For the dissimilar query pair, the queries from different forums are considered. In this way, a dataset of 500 queries with similar, dissimilar pairs was constructed, which is referred to herein as Breast Cancer-Query Query pair (BC-QQP) data set. A sample is shown in Table 2.

For the transfer learning technique in query similarity model following dataset were used:

- The MedQuAD dataset [1], consists of 47,457 question-answer pairs constructed from 12 trusted resources such as NCI (National Cancer Institute), MedlinePlus Health Topics, MedlinePlus Drug etc.
- WebMD dataset [29], which is an online publisher of medical information including articles, videos, and frequently asked questions (FAQ). By using a publicly available crawl over the FAQ of WebMD 46,872 question-answer pairs were extracted.

3.2 Architecture of the proposed system

The architecture of the proposed answer recommendation system is shown in Fig. 1. The first phase of the system is to find a similar archived query for each of the user's new query from the archived query-answer pair. A query similarity model-a Siamese network model, which is very efficient for finding the similarity between text, is used during this phase. During the second phase a summary is generated by summarization process and archived

Table 2 A sample query and its similar and dissimilar pairs

Sample query	Similar query	Dissimilar query
does anyone have any experience with any drugs that help with the neuropathy or even any words of encouragement?	posting for a friend. a dear friend who is going through chemo has been left with mega neuropathy in his feet. Any thoughts?	does anyone else feel that on the masectomy side things feels a bit different/bonyier than on there normal side?
rad onc says no tamoxifen during rads- says increases side effects of rads. medical onc says start 1 month out from surgery. who wins? anyone had issues with tamoxifen during rads?	i waited for tamoxifen till i finished rads. i stopped rads on a friday and started tamox on sunday. what are you comfortable with? Have you taken during rads?	my question is for those that have been through it and are currently using or have used wigs what type do you suggest?

in a query-answer pair repository. The summarization process includes an extractive summarization technique with optimization in the threaded posts. Finally, in the third phase, the corresponding answer is fetched and recommended to the user. Each of these phases is elaborated upon in Sections 3.3, 3.4, and 3.4.4 respectively.

The overall problem can be stated as follows: Let a new query from a user be q_{new} and summary repository contain a set $R = \{(Q, S)\}$, where $\{(Q, S)\} = \{(q_1, s_1), (q_2, s_2) \dots (q_n, s_n)\}$ of n query-summary pairs and let $Q = q_1, q_2, \dots q_n$ be n queries and $S = s_1, s_2 \dots s_n$ be n corresponding summaries. If q_{new} is similar to any of query in query-answer pair repository, $q_{new} \approx q_j$ where $q_j \in q_n$, then corresponding summary $s_j \in s_n$, be recommended to the user.
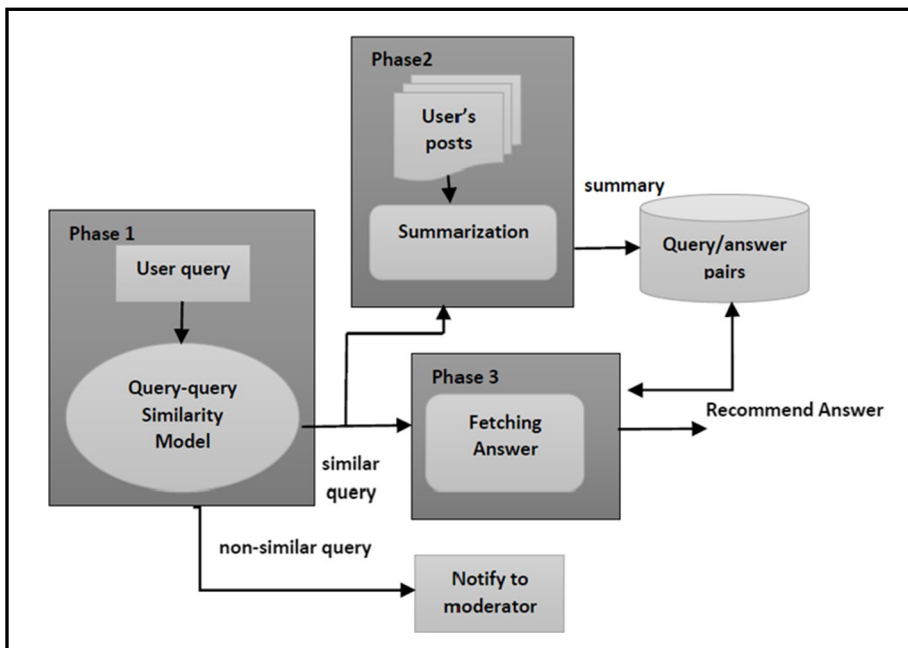

Fig. 1 Proposed Architecture of Recommender System

3.3 Query similarity model using Siamese network

The query similarity model is used to determine the semantic relatedness between two medical queries. As stated in the Related Studies section, the concept of transfer learning enabled us to accomplish this task. A double fine-tuning approach from the transfer learning method was used during this phase [44]. The model was first fine-tuned using a large corpus of medical question-answering dataset and then fine-tuned using our small set of labeled question-question paired dataset. The objective is to integrate medical information into the model so that it can interpret the semantics of each question. First, the Siamese architecture is described and then it is followed by the transfer-learning technique.

3.3.1 Siamese architecture

A Siamese network classifies the query as similar or dissimilar. Figure 2 shows the architecture of Siamese network. The architecture consists of two identical sub-networks. Then, the absolute difference between both the representations is calculated, and similarity score is generated by the final sigmoid layer.

The inputs to the sub-networks are two queries consisting of sequence of words represented by $q_1 = (x_1^1, x_2^1, ...x_n^1)$ and $q_2 = (x_1^2, x_2^2, ...x_m^2)$. Here, q_1 is the current query and q_2 is the archived query, from query-answer pair repository; x_i^1 and x_i^2 are word sequences in first and second query respectively, n and m are number of words. The first layer in the model is the embedding layer, which is a BERT embedding technique. BERT is a bidirectional language understanding model trained on large corpora of English language text. The queries are directly fed to the embedding layer and they are embedded as vectors of 768 dimensions using BERT-Base model. They are then fed to LSTM layer above. To investigate the quality of embedding technique, Bio-BERT was also used in this experiment. Bio-BERT is the general BERT model pre-trained on PubMed abstracts and PMC full-text articles for various
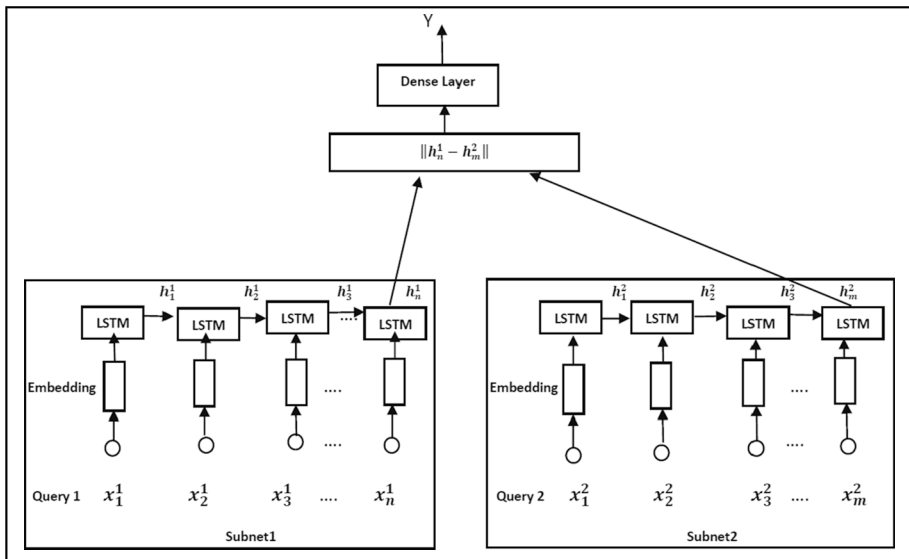

Fig. 2 Query similarity model - Siamese architecture with LSTM

biomedical text mining tasks [22]. In the second layer, LSTM is used to represent the queries from the BERT vectors. LSTM can naturally manage the word order and word sequence. Each LSTM units comprises many hidden layers of gated cells known as memory cells, that can either remember or forget information. For each word x_i, the cell value h_i is computed as a linear combination of current input and the previous state as follows:

$$h_i = f\left(w_h h_{i-1} + w_x x_i\right) \tag{1}$$

Where f is the tanh activation function, w_h is the weight of the hidden layer, h_{i-1} is the previous state, w_x is the weight of the current input, and x_i is the current input word vector. The above calculation is repeated for all hidden states. The final representation of the queries was obtained from the last LSTM units of sub-networks as vectors h_n^1 and h_m^2 respectively. Then, element-wise absolute difference between two vectors is calculated by $\left\|h_n^1 - h_m^2\right\|$ and the value is finally fed to the dense layer where the sigmoid function is used to obtain the similarity score, S, which has the values from 0 to 1. If $S \geq 0.5$, the predicted label Y is 1 (the queries are similar), and if $S < 0.5$, the predicted label Y is 0 (the queries are dissimilar). Binary cross entropy is used as the loss function for each question pair and is defined as follows:

$$Loss = \hat{y}\, log(S) + \left(1 - \hat{y}\right) log(1 - S) \tag{2}$$

Where $\hat{y}$ is the true label and S is the output probability for label 1 and (1-S) is the output probability for label zero. The stochastic Gradient Descent method is used to update the parameters, weights and biases, of sub-networks and is computed by using backpropagation. The implementation was achieved in python using Keras library.

Bidirectional LSTM (BiLSTM) was also used in the sub-networks for the experiments instead of LSTM. Because LSTM is a one-directional approach, it cannot obtain information from the future sequences of words in a sentence. BiLSTM consists of two LSTMs: one for taking input sentence from left to right, and the other for taking input from right to left. Siamese network with BiLSTM is shown in Fig. 3. The forward LSTM is provided with word sequences from left to right i.e., x_1 to x_n and backward LSTM is fed with word sequences from right to left i.e., x_n to x_1. Thus, past data is extracted from first LSTM and future data is extracted from second LSTM. As shown in Fig. 3, the final output h_n^1 for first sub-network is obtained by concatenating two outputs from the forward LSTM h_n^{1f} and backward LSTM, h_n^{1b} using Eq. (3). The same procedure is repeated for second sub-network and the final output h_m^2 is obtained through Eq. (4). The final similarity score S and the label Y are calculated as similar to LSTM network explained above.

$$h_n^1 = h_n^{1f}\left\|h_n^{1b}\right. \tag{3}$$

$$h_m^2 = h_m^{2f}\left\|h_m^{2b}\right. \tag{4}$$

3.3.2 Transfer learning

Transfer learning is implemented as follows. The model is pretrained with MedQuad dataset, which is a large medical question-answer dataset. The objective is to predict the correct answer for a given query. The final tuning is performed with our 500 query-query pair data set (BC-QQP). The queries were highly beneficial for patients undergoing a treatment process.
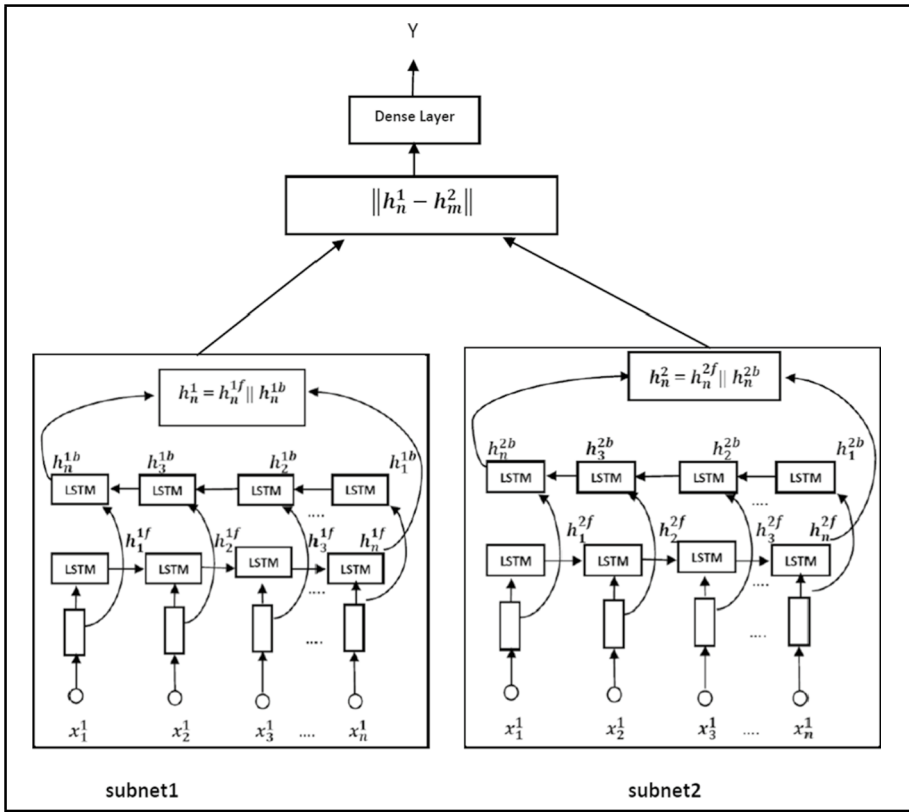
Fig. 3 Query similarity model - Siamese architecture with Bi-LSTM

The tuning methods considered the sentence similarity with a binary cross-entropy loss function, as explained in the Siamese network architecture. For experimental purposes, the model was pretrained using the WebMD dataset and finally tuned with the BC-QQP dataset. The results of both the experiments are tabulated in the Experiment and Result section.

3.4 Summary generation

The second phase of our proposed system was summary generation. The answer to a similar query is found from the query-answer pair repository. To accomplish this goal, the answers from the already discussed posts are summarized and stored. In OHCs, it is common for a user to initiate a thread by asking a question, with several experienced users providing different suggestions and opinions based on their experience. The answered posts should be summarized to extract accurate information. As stated in the introduction, the current study used an extractive summarization technique. In this context, summarization can be considered as a multi-document summarization where the user's answered posts are made up of multiple documents. However, individual posts are much smaller than a stand-alone document, which is typically composed of four to five sentences. Hence, in this work, without losing the context, a summary is created using coherent sentences from the posts.

Let $Q = q_1, q_2, \ldots q_n$ be n queries in threads and $P = \{p_1, p_2, \ldots p_m\}$ be a set of m reply posts corresponding to each query q_j. Each post p_i consists of a set of k sentences, $p_i = \{s_1, s_2, \ldots s_k\}$. A summary is generated as $S = \{s_j^i\}$, where $1 \leq i \leq m$ and $1 \leq j \leq k$, consists of subset of j sentences from m posts. The summary consists of a meaningful subset of sentences that should be coherent and non-redundant. The full summary generation process is comprised three main steps: pre-processing, sentence ranking, and summary generation as depicted in Fig. 4 and described as follows:

3.4.1 Pre-processing

The first step involved pre-processing the sentences from the posts. This step consisted of sentence segmentation, tokenization, stop-word removal and stemming. In sentence segmentation, each post p_i was divided into individual sentences as $s_1, s_2, \ldots s_m$. Each sentence was then tokenized into different words in the second sub-step, as $s_j = \{w_1, w_2 \ldots w_k\}$. From these words insignificant words called stop-words such as 'a', 'the', 'an' etc. were removed. Finally, the words were converted into their base form using the Porter stemming method. All steps were carried out using python's NLTK tool kit.

3.4.2 Sentence ranking

After pre-processing, the sentences were converted into a sentence vector using the BERT technique. Later, the Bio-BERT technique was also used for experimental purpose. The similarity score, sim_score, (cosine similarity), between the sentences in each post were calculated as Eq. (5).

$$sim\_score(s_i, s_j) = \frac{\sum_{i,j=1}^{m} s_i \quad * \quad s_j}{\sqrt{\sum_{1=1}^{m} s_i^2} * \sqrt{\sum_{j=1}^{m} s_j^2}} \tag{5}$$

where s_i and s_j are sentence vectors of the corresponding sentences. For each sentence, the similarity score with other sentences was calculated and then averaged to rank the sentences. The sentences were ranked according to ascending order of the average similarity score followed by selecting the sentences with a score less than a threshold value. In this way, the least similar sentences of each post were selected and merged together to form a single document, D. The sentences were arranged according to the order of posts appearing in the thread.
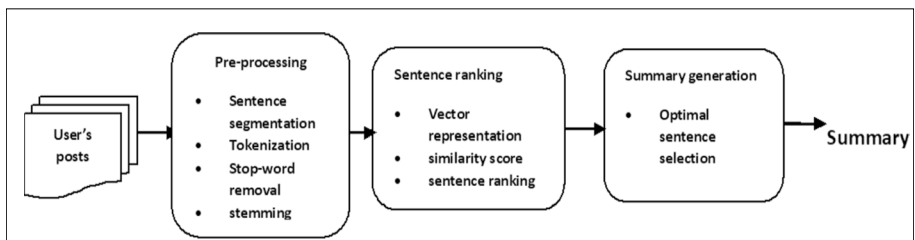
Fig. 4 Summary generation

3.4.3 Summary generation

The next step was to generate more optimal summary from a single document, *D*. The single document from the previous step may contain some similar sentences because it was selected from different posts but discussed a similar topic. Therefore, an optimization technique was applied at this step to generate a more precise and non-redundant summary. An optimization score for each sentence was computed from the three features of the sentences. A good summary should possess three important features: content coverage, non-redundancy and cohesion [32].

Summary, *S* contains a set of sentences s_1, s_2, ...s_n, which should be related to the content of discussed topic. Since this study deals with a query-based summarization, the content coverage was measured based on similarity of the summary with corresponding query q_j, which is termed as *cont_cov*. It was computed using Eq. (6) as follows:

$$cont\_cov(S) = sim\left(s_{avg}, q_j\right) \tag{6}$$

where s_{avg} is the average sentence vector of summary which contains *{s_1, s_2...s_n}* sentences and *sim (s_{avg}, q_j)* is cosine similarity between s_{avg} *and* q_j.

The single document produced from the sentence ranking step contains some redundant sentences. Hence, the dissimilarity among sentences was computed using Eq. (7):

$$non\_redundancy\ (S) = 1 - \max_{s_{i,j \in S}}\left(sim\left(s_i, s_j\right)\right) \tag{7}$$

Cohesion is the conceptual relationship among sentences in the summary; that is, the sentences must discuss same idea of the content [37]. According to [40], cohesion is the ratio of average of all similarities of sentences in the summary to the maximum of similarities. It was computed using Eq. (8):

$$cohesion(S) = \frac{\log\left(Avg_{s_{i \in S}}\left(sim\left(s_i\right)\right) * 9 + 1\right)}{\log\left(Max_{s_{i \in S}}\left(sim\left(s_i\right)\right) * 9 + 1\right)} \tag{8}$$

where $Avg_{s_{i \in S}}\left(sim(s_i)\right)$ is the average similarity of all sentences in the summary, and $Max_{s_{i \in S}}\left(sim\left(s_i\right)\right)$ is the maximum of similarity value among sentences.

Then, an optimization score, *f(S),* was formulated by weighted sum of the above equations as follows:

$$f(S) = \alpha\ (cont\_cov(S)) + \frac{1-\alpha}{2}(non\text{-}redundancy) + \frac{1-\alpha}{2}(cohesion(S)) \tag{9}$$

where the value of α is manually defined between $0 < \alpha < 1$. Here, more weight was provided to the content coverage and equal weight was given to non-redundancy and cohesion. Hence, α was given a value of 0.5 and, weights to the two other components non-redundancy and cohesion were given a value of 0.25 each. The algorithm used to generate a summary is shown in Fig. 5. Sentences with an optimization score, *f(S)≥k* as the threshold value were included in the summary. The generated summary was stored in the repository *R*, as *R={(Q, S)}*.

3.4.4 Answer recommendation

In this phase, the query which is more closely related to the user query is found first. Query q_j to the user's new query, q_{new} was extracted from the query similarity model, $q_{new} \approx q_j$,

Input: Single document $D = \{s_1, s_2, ...s_n\}$ from the sentence ranking step, threshold value k, and query Q
Output: Summary S

1. Select the first sentence, s_1, from D and include it in Summary S
2. For each sentence s_i,
    a. compute the average sentence vector, s_{avg}, with all previous the sentences included in the summary and calculate the content coverage as, $cont\_cov(S) = sim\ (s_{avg},\ Q)$
    b. calculate the cosine similarity using all previous sentences included in the summary and determine the non-redundancy, $non\_redundancy\ (S) = 1$- $\max_{s_i \in S} (\ sim(s_i, s_{i+1}))$
    c. compute the average similarity and maximum similarity with all the previous sentences included in the summary and compute cohesion as, $cohesion(S) = \dfrac{\log\ (Avg_{s_i \in S}(sim(s_i)) * 9 + 1)}{\log(Max_{s_i \in S}(sim(s_i)) * 9 + 1)}$
    d. compute the optimization score, $f(S) = \alpha\ (cont\_cov(S)) + \dfrac{1-\alpha}{2}(non$-$redundancy) + \dfrac{1-\alpha}{2}\ (cohesion(S))$
    e. if $f(S) \geq k$, include s_i in summary S; otherwise, discard the s_i
    f. Repeat a through e until s_n

Fig. 5 Algorithm for summary optimization

where $q_j \in Q$. Then, answer S to the matched query is fetched from query-answer pair repository $R = \{(Q, S)\}$ and recommended to the user.

4 Experiment and results

This section describes the performance evaluation of query similarity model and summarization. We tested and evaluated the performance of twelve query similarity models using BERT and BioBERT embedding techniques as well as transfer learning techniques. We created and evaluated four different summaries using two embedding methods, as well as optimization and non-optimization techniques for summarizing.

4.1 Performance evaluation of query similarity model

The query similarity task is carried out using Siamese network by applying different embedding techniques-BERT and Bio-BERT. The following experiments were conducted to assess the performance of the model. The different word embedding techniques BERT and BioBERT with LSTM and BiLSTM units were implemented to measure the performance. The BC-QQP dataset was directly fed to the Siamese architecture, without any pretraining and the similarity was assessed. The number of LSTM and BiLSTM units was 128 and the number of epochs was 10, with a learning rate of 0.01 and a drop out of 0.2. Eighty percent of data set was used for training and 20% was for testing. The results are listed in Table 3. There were totally 12 models to compare. The base model, model 4, i.e., BiLSTM architecture with Bio-BERT embedding of queries, yielded a better result with an F1-score of 65% than the BERT embedding technique.

To assess the efficacy of transfer learning, the model was pre-trained with MedQuAD and WebMD dataset and the results were compared. Each query-answer pair in those datasets was labeled as positive ('1') and each query with a random another answer was labeled as negative ('0'). These labeled pairs were embedded with BERT technique and fed to each of the subnet of Siamese architecture. The experimented was also conducted with BioBERT technique. Then the model was trained with 10 epochs to classify the dataset as

Table 3 Evaluation of query similarity

Techniques	Models	Precision	Recall	F1-score
Without Transfer learning (Base Model)	BC-QQP + BERT[a] + LSTM (Model 1)	0.48	0.52	0.50
	BC-QQP + BERT + BiLSTM (Model 2)	0.53	0.50	0.51
	BC-QQP + BioBERT[b] + LSTM (Model 3)	0.58	0.61	0.60
	BC-QQP+ BioBERT + BiLSTM (Model 4)	0.63	0.66	0.65
With Transfer learning (MedQuAD)	BC-QQP + BERT + LSTM (Model 5)	0.78	0.72	0.75
	BC-QQP + BERT + BiLSTM (Model 6)	0.81	0.83	0.821
	BC-QQP + BioBERT + LSTM (Model 7)	0.84	0.81	0.83
	BC-QQP+ BioBERT + BiLSTM (Model 8)	0.85	0.86	**0.855**
With Transfer learning (WebMD)	BC-QQP + BERT + LSTM (Model 9)	0.80	0.81	0.805
	BC-QQP + BERT + BiLSTM (Model 10)	0.81	0.82	0.815
	BC-QQP + BioBERT + LSTM (Model 11)	0.83	0.81	0.82
	BC-QQP+ BioBERT + BiLSTM (Model 12)	0.84	0.81	0.824

[a]https://github.com/google-research/bert, [b]https://github.com/dmis-lab/biobert

either '1' or '0'. After this pre-training, 80% of BC-QQP training data set was fed to model for fine-tuning. Finally, the model was tested with 20% of test data. A more promising result was obtained with the model, model 8, that was pre-trained using MedQuAD data set, achieving an F1-score of 85.5% (value is highlighted in bold in the Table 3), as compared to WebMD dataset. An improvement in the F1-score of more than 20% was obtained from the base model, model 4. Training loss for 10 epochs is also plotted for both the LSTM(model 7) and BiLSTM(model 8) models, as shown in Fig. 6. The BiLSTM model reduces training loss more effectively than the LSTM model.

4.2 Performance evaluation of summarization

BC-QA dataset consists of 500 queries and its replies. After pre-processing, the sentence vectors for each of the sentence in the reply posts were built by utilizing BERT technique. Then, average cosine similarity of each sentence in the posts was computed and the sentences with an average score of less than 0.5 were chosen for constructing the single document, from which a summary was generated by applying the optimization algorithm. The optimal sentences for summary were selected by choosing a threshold value k as 0.7. To assess the quality of sentence embedding in summary generation, the Bio-BERT technique was too used to generate sentence embedding.

Fig. 6 Comparison of LSTM and BiLSTM models training loss
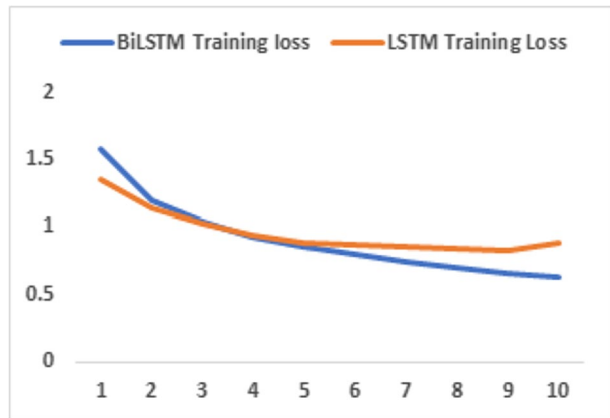
An automatic summary evaluation metric called Recall-oriented Understudy for Gisting Evaluation (ROUGE) [24] was used to test the quality of generated summary. The ROUGE score is calculated based on intersection of N-gram between system-generated summary and manual summary. The score is calculated based on the following equation [35]:

$$ROUGE\text{-}N = \frac{\sum_{s \in manual\ summary} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{s \in manual\ summary} \sum_{gram_n \in S} Count(gram_n)} \tag{10}$$

Where $gram_n$ is the length of n-gram in summary, $count_{match}(garm_n)$ is the number of n-gram matches between the manual summary and system-generated summary. A sample of system-generated summary and manual summary are shown in Table 4.

Table 4 A sample of System-generated summary and manual summary

Query: i am having bi-lateral diep flap reconstruction. What kind of help do i really need after i get home? do i need someone here all the time the first few days or can i get by with regular check-ins?

Manual Summary	System-generated Summary
everyone is different on how they react to surgery but i found i was quite capable of being by myself during the day. i also got an inexpensive shower seat which really helped for a couple weeks. your not allowed to raise your arm higher than shoulder height so washing my hair sitting down was much easier. prepare your house in advance. put everything within reach both bending and stretching will be difficult. having help your first day or two is good if possible someone to wait on you bring you food & drinks and help you get out of bed but if you have to do it alone you probably could if you prepared meals in advance and had everything you need within reach including meds.	foot girl try reading a few pages from diep 2015 everyone is different on how they react to surgery but i found i was quite capable of being by myself during the day. i also got an inexpensive shower seat which really helped for a couple weeks. your not allowed to raise your arm higher than shoulder height so washing my hair sitting down was much easier. prepare your house in advance. put everything within reach both bending and stretching will be difficult. having help your first day or two is good if possible someone to wait on you bring you food & drinks and help you get out of bed but if you have to do it alone you probably could if you prepared meals in advance and had everything you need within reach including meds. lots of pillows help with getting comfortable. it is hard in the beginning you are very tired and weak. have someone there when you shower just incase.

Table 5 ROUGE-1 score of summaries generated

Summary#		ROUGE-1 score
1	System generated summary- BERT (without optimization)-Baseline	0.386
2	System generated summary- Bio-BERT (without optimization)	0.412
3	System generated summary- BERT (with optimization)	0.433
4	System generated summary- Bio-BERT (with optimization)	0.491

Separate score from ROUGE can be obtained, when n = 1,2,3,4-g matching. Among these scores, more agreement with manual summary was obtained with n = 1, 2 that is, unigram and bigram-based scores. Both the score values were tabulated in Tables 6 and 7. Conventionally, summarization task was carried out by validating with more than one human written summary. More than one team may write the summary, check the agreement with both the team and validate the system generated summary. But with the peculiarity of more medical content in our posts, the summary produced was done by only one clinical team in our study.

To evaluate the effectiveness of our proposed summarization technique, four summaries were generated based on BERT and BioBERT vectorization techniques with and without optimization step. The summary generated after sentence ranking step was considered as summary without optimization. Two different ROUGE scores were compared corresponding to unigram and bigram length of words. Tables 5 and 6 shows Rouge-1 and Rouge-2 scores respectively. Here also, the Bio-BERT embedding technique showed better result over BERT. Note that, with optimization step, Rouge score was significantly improved by a factor of ≈8%. Thus, a Rouge score of 49.1% was obtained using Bio-BERT with optimization step for unigram words. For bi-gram words, a Rouge score of 26.7% was obtained with the same approach.

Precision, Recall, F1-score, and Accuracy metrics are also used to assess the four summaries, which are based on retrieved correct sentences in the system generated summary. They are defined as follows:

1. Precision: is the ratio of number of retrieved correct sentences to the total of retrieved correct sentences and retrieved incorrect sentences in the summary.
2. Recall: is the ratio of number of retrieved correct sentences to the total of retrieved and non-retrieved correct sentences in the summary.
3. F1-score: is the ratio of harmonic between Precision and Recall.
4. Accuracy: is the ratio of total of retrieved correct and non-retrieved incorrect sentences to the total sentences in the summary.

Table 6 ROUGE-2 score of summaries generated

Summary #		ROUGE-2 score
1	System generated summary- BERT (without optimization)-Baseline	0.198
2	System generated summary- Bio-BERT (without optimization)	0.213
3	System generated summary- BERT (with optimization)	0.235
4	System generated summary- Bio-BERT (with optimization)	0.267

The graphical depiction of these metrics in respect to four summaries is shown in Fig. 7. The Bio-BERT with optimization technique is found to have the maximum accuracy of 82%. Precision, Recall, and F1-scores for the same summary are all quite high when compared to other summaries. The efficiency of the optimization technique is again demonstrated by these outcomes.

4.3 Qualitative analysis

To get a better insight into our query similarity model and summarization technique, we performed error analysis. Table 7 shows some instances of query pairs with their actual labels and the labels from the four query similarity models, model 4, model 8 and models 12. The reason for considering these models was the fact that Bio-BERT with Bi-LSTM architecture was better than other architectures. It was noted that all models could understand the query pairs 2 well, but were unable to do so with query pairs 3. In pairs 3, though the question was about the position during rad, the terms 'arm position' and 'rads position' as well as the negation part ('but doesn't look likely') could be the reason for giving wrong result from all the models. However, the model 8 provided the exact result for all other query pairs.

In order to clarify the efficiency of the best model, Model 8, in understanding the query pairs, further analysis was carried out. Table 8 demonstrates an analysis of this kind, in which a query was combined with other pairs of slight tweaks. The query pair 3 from Table 7 is taken and updated repeatedly before the model correctly labels it. This sample shows the point at which the model predicts the output correctly and also highlight the need for additional training data. This kind of improved query pairs were not considered for quantitative analysis and these are used for understanding the best model. The above
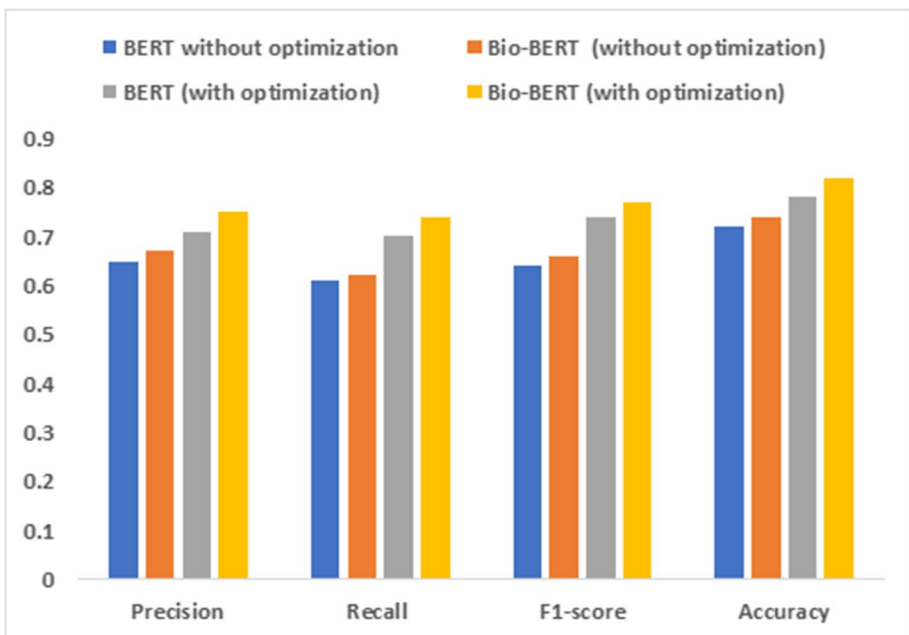
Fig. 7 Comparison of Precision, Recall, F1-score and Accuracy of four summaries

Table 7 Sample query pairs with actual and predicted labels

	Query 1	Query 2	Actual Label	Labels from models		
				Model 4	Model 8	Model 12
1	i am wondering if anyone was very tired already after a couple of rads?	anyone else feel this tired from almost the start of rads? did it get worse or get better?	similar	dissimilar	similar	dissimilar
2	does anyone else feel that on the mastectomy side things feels a bit different/bonyier than on there normal side?	My surgery side is very bonyier than the normal side. Has any one get similar experience?	similar	Similar	similar	Similar
3	wanted to see if anyone else got "permission" to alter the arm position during rads?	i'm praying that they might be willing to consider other options for rad position but doesn't look likely. Any experience?	Similar	dissimilar	dissimilar	Dissimilar
4	the skin between my breasts on the radiated side is covered with a red rash that itches. my nipple is dark brown with dark dots all around it. is this normal?	Even after my rads my breast is inflammated and red. Is this normal?	Dissimilar	Similar	dissimilar	Similar

Table 8 Sample query pairs with different changes in same pairs

	Query 1	Query 2	Actual Label	Labels from model 8
1	wanted to see if anyone else got "permission" to alter the arm position during rads?	i'm praying that they might be willing to consider other options for rad position but doesn't look likely. Any experience?	Similar	dissimilar
2	wanted to see if anyone else got "permission" to alter the arm position during rads?	Are they willing to consider other positions for rad, but doesn't look likely? Any experience?	Similar	dissimilar
3	wanted to see if anyone else got "permission" to alter the arm position during rads?	Are they willing to consider other arm positions for rad, but doesn't look likely? Any experience?	Similar	dissimilar
4	wanted to see if anyone else got "permission" to alter the arm position during rads?	Are they willing to consider other arm positions for rad? Any experience?	Similar	similar

two analyses provide us with a clear insight into the interpretability of our model and the need for additional training data.

Table 9 shows the qualitative analysis of our proposed summarization technique, with optimization and without optimization, by Bio-BERT embedding. It is noted that without optimization, the second and third lines convey an almost similar meaning. But with the optimization technique, the second sentence has been removed, and the third sentence, which is closer to the query, remains. Compared to the manual summary, this sentence (text in italics - "i finished a year and half ago…") is an additional one, even though it conveys detailed information. Another point to mention here is that there is an extra portion of the sentences in summary. For instance, "i had a doc tell…" (text in italics) in the first line and "i bet" in the last sentence are additional parts in the summary with optimization. That may be attributed to the fact that the algorithm ultimately treats the sentences as a whole, not any part of it. But it can be well handled at the time of manual formation.

5 Discussion

5.1 Principal findings

Answer recommendation is extremely popular in community forum sites. However, in health community forums, its applicability has not been extensively studied. A few experiments have been conducted on medical community forums to recommend answers, and they are all focused on the best answer depending on voting system. In most health community forums, despite survivors or expert patients share their experiences, no rating for posts is given. For instance, if a patient expresses a concern or query about a side effect, treatment procedure or a particular situation, others share their experiences, suggestions, positive or negative feedbacks. In this context, the current study is extremely significant in providing a precise answer considering all the responses shared by experienced patients responded to the user's concern. More samples of system-generated summary with corresponding queries and similar queries are shown in Appendix.

For several reasons, patients are reluctant to share their concerns with health care providers. This may be due to embarrassment, or not even being conscious of the existence of the problem. An example of such a situation is shown Table 10. Hence, in this context, an automated patient-centric recommendation system is a way to enable users to obtain advice or answers addressing their concerns.

From the performance evaluation of the study, the query similarity model with transfer learning is very effective in capturing the medical knowledge in our data set. Since the BC-QQP data set is specifically for breast cancer patients, it contains more drug related and cancer related terms. With the MedQuAD data set, the model can generalize and learn the medical queries well. This is because MedQuAD contains more question-answer pairs that are related to cancer and cancer related drugs than the WebMD dataset. When we compared the findings of our proposed method to some of the studies listed in the related studies, we found that our method is effective. A very recent study in [25] found that similar queries related to COVID-19 achieved an accuracy of 84.2%. Their approach was based on language model BERT and contained more than 3000 COVID-19 query pairs. But with our Siamese network and Bio-BERT embedding technique, our model resulted in a more promising result of 85.5%. When we compared our results to [41], we found that our model outperformed. Their precision in the first 20 results was 86%, whereas our model's

Table 9 Comparison of summary without optimization and with optimization

Query: does anyone else feel that on the mastectomy side things feels a bit different/bonyier than on there normal side?

Answers	Actual Summary	System generated Summary (without optimization)	System generated Summary (with optimization)
1. i had a doc tell me last week that it can take up to 3 years for the surgical area to stop changing. i thought that was an awfully long time but i guess that was her way of telling me i'm going to continue to feel pains and swelling changes in that area while the nerves continue to rebuild. 2. it doesn't sound like it's anything to be concerned about!. i finished a year and a half ago and i still feel soreness/tenderness all around that area including collar bone and below/above it. i bet some inflammation around it is completely normal.	it can take up to 3 years for the surgical area to stop changing. some inflammation around it is completely normal.	i had a doc tell me last week that it can take up to 3 years for the surgical area to stop changing. i thought that was an awfully long time but i guess that was her way of telling me i'm going to continue to feel pains and swelling changes in that area while the nerves continue to rebuild. i finished a year and a half ago and i still feel soreness/tenderness all around that area including collar bone and below/above it. i bet some inflammation around it is completely normal.	*i had a doc tell me last week that* it can take up to 3 years for the surgical area to stop changing. *i finished a year and a half ago and i still feel soreness/tenderness all around that area including collar bone and below/above it. i bet some* inflammation around it is completely normal

precision in the first 20 results was 89%. The summary generated with optimization also increased the efficacy of the summarization phase. The optimization score for generated summary has achieved by varying the tuning parameter k which was ideal when $k \geq 0.70$. When comparing our summarization approach to that of Rautray and Balabantaray [32], the results show a substantial improvement. Their best approach only provided a Rouge-1 score of 0.43, but our technique produced a Rouge-1 score of 0.49.

6 Conclusion and future direction

This is a pioneering study in the area of breast cancer intended to help the patients make quick and informed decisions. The study demonstrated a system for recommending an answer to a particular query related to breast cancer. The recommended answer is the summary of a discussion in the forum by experienced patients, moderators of forums and survivors. These answers contain experiences, suggestions, solutions, viewpoints and above all, how they managed a particular situation. The study contains mainly three phases: similar query retrieval, summary generation and answer recommendation. The Siamese network applied was very successful in finding the similar queries. The Bi-LSTM with Bio-BERT embedding technique outperformed all other models and provided f1-score of 85.5%, a promising performance than many previous studies. The optimization approach utilized was also highly promising in creating the summary, as seen by Rouge1-score, 0.49, during the summarizing phase. The query-answer pairs created in the study offer a great insight into the multiple challenges that patients encounter during the treatment of breast cancer. This information can help clinicians to address the issues they encounter outside clinical matters.

The study was limited to 500 query-answer pairs on the area of breast cancer area and we therefore intend to include a greater number of queries and its specific answers in the future. Although the question similarity model worked well, there were still a few cases in which our model was unable to find a similarity correctly. For instance, two queries – *"did you have more work done after the initial diep surgery? i'm wondering if i should expect a lot of swelling in my abdomen from the liposuction?"* and *"i am having bi-lateral diep flap reconstruction. what kind of help do i really need after i get home?"* were predicted to be similar, despite being different. Both queries were about diep flap reconstruction surgery. The first one was focused on post-surgery complications and the second deals with how to manage the post-surgery state. This indicates that the model requires more tuning to handle queries that are closely related but conceptually distinct and to generalize the similarities. Future work should take care of such concerns. The present study can be extended to other health forums to obtain awareness in other areas of concerns.

Table 10 A Sample query and its answer

Query	Answer
"tell me about it I am so confused and I do not want to ask my dr it will just confuse me more. dcis was not even mentioned to me at the beginning, and yet it's there somehow, it's there the scans do not show everything. feeling very confused and fed up. Any advice?"	dcis is often found with idc. they are not treating your dcis just your idc. It cannot respond to systemic treatments such as chemo, so the size of your dcis would not change from having chemo. sometimes after a biopsy or lumpectomy/mastectomy, they found more (or less) than they originally thought was there initially. I think a second opinion is a good idea in either case.

Appendix

A sample of 10 queries, similar query and its answer

Query	Similar query	Answer (system-generated summary)
has anyone had the port placed in the arm (inner side of arm a few inches under the arm pit)? Do you have any problem?	can you work out and do physical things with port placement with no problem?	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. if it means saving your veins the trauma i would recommend going with the port. no general for the insertion and when i had it removed it was right in the doctors office. it's took me about a week to get use too it. my port is inside my left arm just above my elbow. as few chemos as you're getting and no herceptin i'd probably go with no port as well. but again in your situation if your veins are good i don't see a reason to do the extra surgery.
several remember their chemo being red. are they all red?	i've been listening to people tell their cancer survival stories since my diagnosis. they no longer can even drink anything red as a result. Are there anything red chemo?	i had adriamycin and as people have told you it's red. i knew it was called the red devil. it didn't bother me and as for sharon it didn't stopped me to like red wine. if you get the red devil as it's called you will also pee red for a few hours. i drank water laced with cranapple juice. i drank so much water during the infusion that by 2 hours later i had peed all the red out again. and like others it never occured to me to feel bad about red drinks afterwards.
hello i think i have thrush??? what do i do for it???	I have thrush coats in my tongue and inside of mouth. What could be done?	if you think you have thrush you need to tell your doctor. only way to treat it is with oral antifungals. i can tell you that when i was on taxol i had terrible mouth sores and a coated mouth but it was not thrush. thrush coats your tongue and inside of your mouth with a white looking fungus. if it is also mouth sores try salt water solution and if it hurts anything that makes canker sores etc. in addition to salt 1/2 teaspoon baking soda may be added to the saline solution. i had thrush the doctor prescribe a mouth rinse and swallow of niastatin. it took about 3 weeks to go away. also once cleared up stay away from sugary foods and eat plain yogurt. your onc can prescribe diflucan. keep away from yeasty things for now: breads wine or beer etc.

Query	Similar query	Answer (system-generated summary)
does anyone have any experience with any drugs that help with the neuropathy or even any words of encouragement?	posting for a friend. a dear friend who is dealing with lung cancer has been left with mega neuropathy in his feet. Any thoughts?	i know some people are helped by taking gabapentin. you'll have to find out the drug name online. your friend should know that exercising makes it better so does soaking your feet in warm water with epsom salts. the gabapentin is Neurontin the later med pregabalinor lyrica is much much better for my neuropathy. it's important to keep the muscles strong in the legs and feet for balance. drugs for numbness won't bring back feeling in my understanding but they can help with annoying tingling sensations or pain. hopefully your friend's neuropathy which is chemo-induced will lessen and maybe disappear. please tell him to check the bottom of his feet every day to make sure he hasn't injured them or stepped on anything sharp and always wear a soled shoe/slipper.
i am not drenched but my forehead is a little sweaty. is this normal from hot flashes? and what are you ladies doing to control it?	hi- my last chemo was nov 22nd (tch) and ever since ive been getting hot flashes day and at night its less. What is your suggestion for this?	doctor prescribed effexor 37.5mg it took about 6 weeks to build up in my system but after that i was 95% hot flash free. if you can get your wrists cool you can cool off your whole body. apply that cold soda to the sides of your neck too. i am taking a low dose blood pressure med and it helps to keep mine from being so severe. i have found that some foods trigger mine. especially salty or sweet foods.
hello everyone:i had a lumpectomy in aug 09 and a cancer of 1.3mm - stage 0. next week i am scheduled to have a planning session with a ct scan and measurements before radiation. can someone tell me why i need a ct scan?	My rads onc told me take ct scan before rads. Has any one got suggestions like this?	i had my ct scan last week was told it was so they know the position of my heart and lungs before they start any radiation treatment. they do a ct scan so they can put on the little dots and everyone knows where to line up the machines for radiation. the tattoos are no more than little pin pricks and i don't think mine bled at all. they are so tiny it is hard to fine them later on.
rad onc says no tamoxifen during rads- says increases side effects of rads. medical onc says start 1 month out from surgery. who wins? anyone had issues with tamoxifen during rads?thanks	i waited for tamoxifen till i finished rads. i stopped rads on a friday and started tamox on sunday. what are you comfortable with? Have you taken during rads?	my dr told me to wait a full 2 weeks after radiation before i start tamoxifen. rad onc said he prefers after so you don't have the side effects mixed up - not knowing what is causing what. i started mine around the same time i did radiation and i did just fine. didn't have any side effects from the radiation. i started tamoxifen as soon as chemo ended and took it all through radiation. i had very little reaction to the radiation

Query	Similar query	Answer (system-generated summary)
the surgeon basically gave me 3'options: lumpectomy & chemo , single mastectomy , double mastectomy & chemo to put my chances of any recurrences extremely low. ladies i'd love an opinion on what you guys decided with your surgeries. Also want to know about reconstruction.	lumpectomy & chemo, single mastectomy or double mastectomy & chemo which one reduce reccurence? What about reconstruction?	excellent post about deciding between lumpectomy and mastectomy : https://community.breastcancer.org/forum/91/topic/. the choices of lx mx bmx i will share that i approached the decision by sharing with the bs the rad doc and the mo my personal priorities in life and work along with treatment. and like you i do have a family history of breast cancer. all of these specialists met in what is called tumor board and make recommendations for treatment. i had no personal preference for lx or mx other than not wanting unnecessarily aggressive surgery. i chose against immediate reconstruction deciding that i wanted to heal and experience life without breasts and pursue a choice of the many reconstruction methods available only after i knew i was dissatisfied with life without reconstruction. i hate the idea of losing my breasts but not so much as to further risk my life so if it is a high likelihood that it will return i think bmx might be the right choice for me.
i have just been diagnosed with breast cancer stage 1 .i have opted for a bi-lateral mastectomy and reconstruction. i guess what i need as far as advise which may be very trivial to most but has been a discussion right now in my house is to get or not to get nipple reconstruction?	Whats your opinion about nipple reconstruction for a bi-lateral mastectomy with stage1 breast cancer?	i hope that you know that no reconstructed nipple works sexually. my ps said they really look good so i'm going to be getting them. i had diep recon and they do feel like real breasts so hoping the nipples will be a nice finishing touch. i had nipple-sparing bmx and kept my original ones. but they are numb. However there might be some hope for us. my plastic surgeon is amazing and he did some fat grafting for me--injectable fat from our own bodies to build up/shape our new breasts. i opted for areola and nipple reconstruction topped off with tattoos by a tattoo artist
it seems a lot of decisions will be decided on which treatment plan is required. but i wonder how will we know if chemo or radiation is required? if i avoid rads he is suggesting going straight to implants. so anyone else go straight to implants?	Can you suuggest me about decision plan about chemo or radiation? What do you think of straight to implant by avoiding rads?	if you do end up having radiation some doctors feel that radiated skin rejects implants. Usually, you have surgery and the doctor removes your tumor. the exact size of the tumor isn't revealed until pathology looks at it after surgery. after surgery your doctor will call you in and give you your pathology report. to my knowledge if the cancer has spread to any of the lymph nodes the doctor will recommend chemo and radiation. the chemo is for any straggling cancer cells that might have gotten past your lymph nodes to prevent distant recurrence. radiation is done to the chest area and that's to prevent local recurrence. chemo is usually done before radiation. some take the oncotype test to determine the benefits of chemo. but for now concentrate on being positive

Data availability The data that support the findings of this study are available from Breastcancer.org but restrictions apply to the availability of these data. Data are however available from the authors upon reasonable request and with permission of Breastcancer.org.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

1. Abacha AB, Demner-Fushman D (2019) A question-entailment approach to question answering. BMC Bioinform 20(1):511
2. Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–749
3. Alahmari N, Alswedani S, Alzahrani A, Katib I, Albeshri A, Mehmood R (2022) Musawah: a data-driven AI approach and tool to co-create healthcare services with a case study on cancer disease in Saudi Arabia. Sustainability. 14(6):3313
4. Arora S, Liang Y, Ma T (2017) A simple but tough-to-beat baseline for sentence embeddings. In: ICLR 2017
5. Asgari H, Masoumi B, Sheijani OS (2014) Automatic text summarization based on multi-agent particle swarm optimization. In: Intelligent systems (ICIS). 2014 Iranian conference on, IEEE, February, pp 1–5
6. Badry RM, Eldin AS, Elzanfally DS (2013) Text summarization within the latent semantic analysis framework: comparative study. Int J Comput Appl 81(11)
7. Balakrishnan A, Idicula SM, Jones J (2021) Deep learning based analysis of sentiment dynamics in online cancer community forums: an experience. Health Informatics Journal 27(2):14604582211007537
8. Baumel T, Eyal M, Elhadad M (2017) Query focused abstractive summarization: incorporating query relevance, multi-document coverage, and summary lengthconstraints into seq2seq models. arXiv preprint arXiv:1801.07704
9. Bhatia S, Biyani P, Mitra P (2014) Summarizing online forum discussions–can dialog acts of individual messages help? Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)
10. Bogdanova D, dos Santos C, Barbosa L, Zadrozny B (2015) Detecting semantically equivalent questions in online user forums. In: Proceedings of the nineteenth conference on computational natural language learning. Association for Computational Linguistics, Beijing, pp 123–113. https://doi.org/10.18653/v1/K15-1013
11. Chali Y, Islam R (2018) Question-question similarity in online forums. In: Proceedings of the 10th annual meeting of the forum for information retrieval evaluation (FIREâĂŹ18). Association for Computing Machinery, 21âĂŞ28
12. Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: pre-training of deep bidirectional transformers for language understanding arXiv preprint arXiv:1810.04805
13. Elsweiler D, Harvey M, Ludwig B, Said A (2015) Bringing the healthy into food recommenders. DRMS Workshop
14. Fan H, Smith SP, Lederman R, Chang S (2010) Why people trust in online health communities: an integrated approach. In: 21st Australasian conference on information systems
15. Farrell RG, Danis CM, Ramakrishnan S, Kellogg WA (2012) Intrapersonal retrospective recommendation: lifestyle change recommendations using stable patterns of personal behavior. In: Proceedings of the First International Workshop on Recommendation Technologies for Lifestyle Change (LIFE-STYLE 2012), Dublin, Ireland, Citeseer, pp 24.
16. Gupta R, Béchara H, El Maarouf I, Orasan C (2014) Uow: Nlp technique developed at the university of Wolverhampton for semantic similarity and textual entailment. In: SemEval@COLING
17. He YX, Liu DX, Ji DH, Yang H, Teng C (2006) Msbga: a multi-document summarization system based on genetic algorithm. In: Machine Learning and Cybernetics, 2006 International Conference on, IEEE, pp 2659–2664
18. Jones J et al (2018) Novel approach to cluster patient-generated data into actionable topics: case study of a web-based breast cancer forum. JMIR Med Inform 6(4):e45
19. Karwa S, Chatterjee N (2014) Discrete differential evolution for text summarization. In: Information technology (ICIT). 2014 international conference on. IEEE, pp 129–133
20. Kiros R, Zhu Y, Salakhutdinov RR, Zemel RS, Torralba A, Urtasun R, Fidler S (2015) Skip-thought vectors. In: NIPS

21. Le QV, Mikolov T (2014) Distributed representations of sentences and documents. In: ICML
22. Lee J et al (2019) BioBERT: pre-trained biomedical language representation model for biomedical text mining. arXiv preprint arXiv:1901.08746
23. Li M, Shi J, Chen Y (2019) Analyzing patient decision making in online health communities. 2019 IEEE International Conference on Healthcare Informatics (ICHI). IEEE
24. Lin C-Y, Hovy E (2003) Automatic evaluation of summaries using n-gram co-occurrence. In Proceedings of Language Technology Conference (HLT-NAACL 2003), Edmton, Canada, May 27 - June 1
25. McCreery CH et al (2020) Effective transfer learning for identifying similar questions: matching user questions to COVID-19 FAQs. Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining
26. Miroševič Š et al (2019) Prevalence and factors associated with unmet needs in post-treatment cancer survivors: a systematic review. Eur J Cancer Care 28(3):e13060
27. Mueller J, Thyagarajan A (2016) Siamese recurrent architectures for learning sentence similarity. In: AAAI
28. Nasralah T, Noteboom CB, Wahbeh A, Al-Ramahi MA (2017) Online health recommendation system: a social support perspective
29. Nielsen LR (2017) Medical question answer data. https://github.com/LasseRegin/medical-question-answer-data
30. Pagliardini M, Gupta P, Jaggi M (2018) Unsupervised learning of sentence embeddings using compositional n-gram features. In: NAACL-HLT
31. Rautray R, Balabantaray RC (2015) Comparative study of DE and PSO over document summarization. In: Intelligent computing, communication and devices. Springer India, pp 371–377
32. Rautray R, Balabantaray RC (2018) An evolutionary framework for multi document summarization using cuckoo search approach: MDSCSA. Appl Comput Inform 14(2):134–144
33. Roitman H, Messika Y, Tsimerman Y, Maman Y (2010) Increasing patient safety using explanation-driven personalized content recommendation. In: Proceedings of the 1st ACM international health informatics symposium. ACM, pp 430–434
34. Rokicki M, Herder E, Demidova E (2015) Whats on my plate: towards recommending recipe variations for diabetes patients. Proc. of UMAP 15
35. Sarkar K (2009) Using domain knowledge for text summarization in medical domain. Int J Recent Trends Eng 1(1):200
36. Sezgin E, Ozkan S (2013) A systematic literature review on health recommender systems. In: E-Health and Bioengineering Conference (EHB), 2013. IEEE, pp 1–4
37. Sharreghi E, Hassanabadi LS (2008) Text summarization with harmony search algorithm-based sentence extraction. Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology
38. Tai KS, Socher R, Manning CD (2015) Improved semantic representations from tree-structured long short-term memory networks. In: ACL
39. Verberne S et al (2018) Creating a reference data set for the summarization of discussion forum threads. Lang Resour Eval 52(2):461–483
40. Verma P, Om H (2019) A novel approach for text summarization using optimal combination of sentence scoring methods. Sādhanā 44(5):1–15
41. Wang J et al (2016) An answer recommendation algorithm for medical community question answering systems. 2016 IEEE international conference on service operations and logistics, and informatics (SOLI). IEEE
42. Waterworth S, Luker KA (1990) Reluctant collaborators: do patients want to be involved in decisions concerning care? J Adv Nurs 15(8):971–976
43. Wiesner M, Pfeifer D (2014) Health recommender systems: concepts, requirements, technical basics and challenges. Int J Environ Res Public Health 11(3):2580–2607
44. Xu Y et al (2019) Double transfer at mediqa 2019: multi-source transfer learning for natural language understanding in the medical domain. arXiv preprint arXiv:1906.04382
45. Yang CC, Jiang L (2018) Enriching user experience in online health communities through thread recommendations and heterogeneous information network mining. IEEE Trans Comput Soc Syst 5(4):1049–1060
46. Zhao X, Tang J (2010) Query-focused summarization based on genetic algorithm. In: 2010 international conference on measuring technology and mechatronics automation. IEEE, pp 968–971