Received: 23 November 2021 | Revised: 14 January 2022 | Accepted: 11 February 2022

CAAI Transactions on Intelligence Technology

DOI: 10.1049/cit2.12081

REVIEW

Analysis of community question-answering issues via machine learning and deep learning: State-of-the-art review

Pradeep Kumar Roy[1] | **Sunil Saumya**[2] | **Jyoti Prakash Singh**[3] | **Snehasish Banerjee**[4] | **Adnan Gutub**[5]

[1]Department of Computer Science and Engineering, Indian Institute of Information Technology, Surat, India

[2]Indian Institute of Information Technology, Dharwad, India

[3]Department of Computer Science and Engineering, National Institute of Technology, Patna, India

[4]The York Management School, University of York, York, UK

[5]Computer Engineering Department, College of Computer & Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia

Correspondence
Adnan Gutub, Computer Engineering Department, College of Computer & Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia.
Email: aagutub@uqu.edu.sa

Abstract

Over the last couple of decades, community question-answering sites (CQAs) have been a topic of much academic interest. Scholars have often leveraged traditional machine learning (ML) and deep learning (DL) to explore the ever-growing volume of content that CQAs engender. To clarify the current state of the CQA literature that has used ML and DL, this paper reports a systematic literature review. The goal is to summarise and synthesise the major themes of CQA research related to (i) questions, (ii) answers and (iii) users. The final review included 133 articles. Dominant research themes include question quality, answer quality, and expert identification. In terms of dataset, some of the most widely studied platforms include Yahoo! Answers, Stack Exchange and Stack Overflow. The scope of most articles was confined to just one platform with few cross-platform investigations. Articles with ML outnumber those with DL. Nonetheless, the use of DL in CQA research is on an upward trajectory. A number of research directions are proposed.

KEYWORDS

answer quality, community question answering, deep learning, expert user, machine learning, question quality

1 | INTRODUCTION

Community question-answering sites (CQAs) continue to serve as useful avenues for Internet users to exchange knowledge. CQAs such as Yahoo! answers (YA), Stack Exchange (SE) and Stack Overflow (SO) attract substantial attention from netizens [1]. For example, YA is supported by more than 120 million users [2]. SO, a popular CQA, dedicated to software development is frequented by more than 50 million monthly visitors (as reported online: http://stackoverflow.com/company). On CQAs, users ask questions, provide answers, submit comments, and evaluate posts (questions and answers) created by others [3, 4]. In so doing, they earn rewards or reputation scores that in turn reflect their expertise to the online community [5, 6]. Clearly, CQAs consist of three modules: the question module, the answer module, and the user module, as shown in Figure 1. A questioner posts a question and waits for answers from others. A question usually includes a title, a description, and keywords, describing the question's topic. The question may be routed to specific users based on topics or tags provided with the question [7]. The answerer submits answer to the question. A single question may fetch zero or more answers. The questioner may mark an answer as the best answer, and the question is labelled as resolved and stored in the site repository for others to view. However, the question may still fetch some new answers from other users [8]. Other users further have the option to comment, upvote or downvote questions as well as answers [9].

Over the last couple of decades, CQAs have been a topic of much academic interest. Scholars have often leveraged traditional machine learning (ML) and deep learning (DL) approaches to explore the ever-growing volume of content that CQAs engender. To this end, the focus of this paper is to clarify the accumulated state of knowledge in the CQA

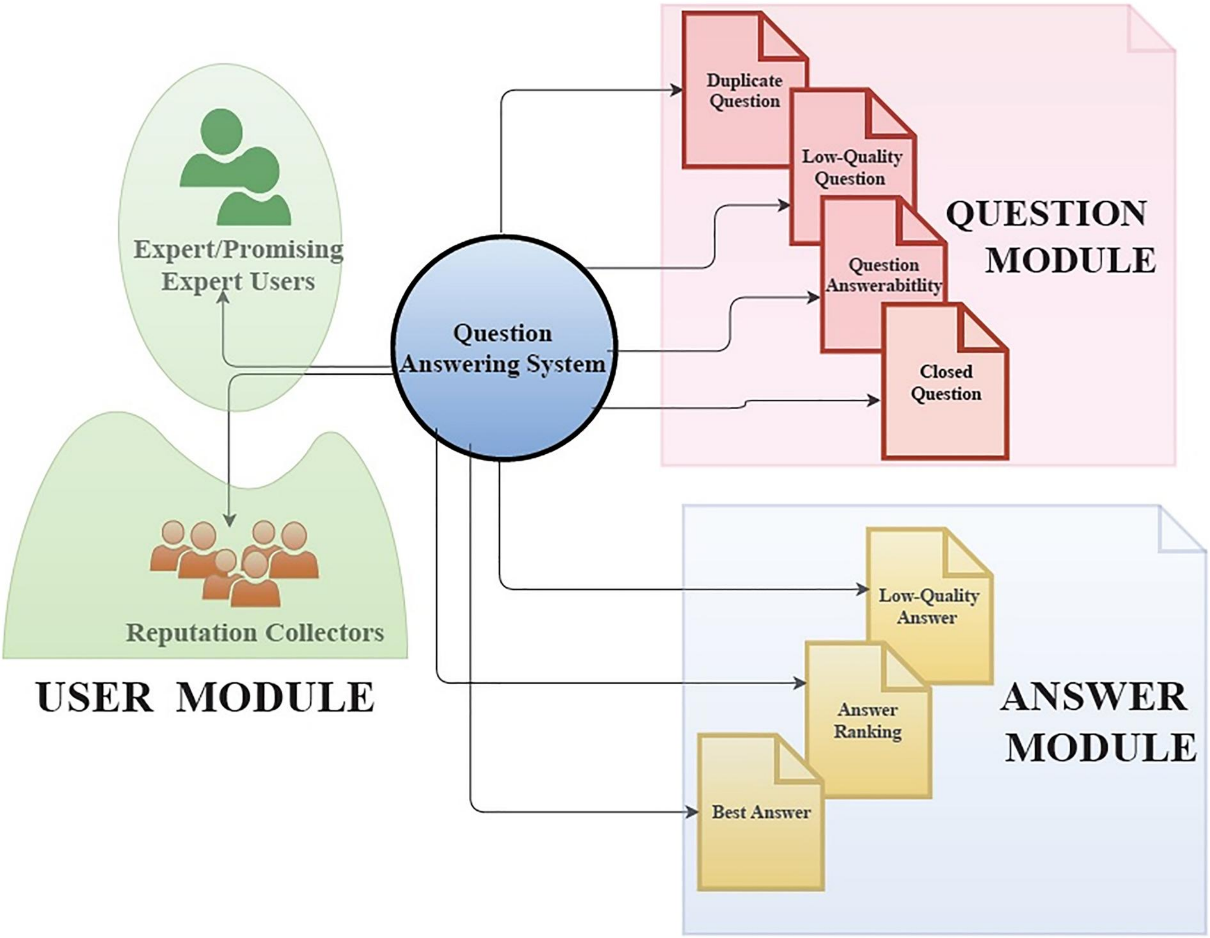

FIGURE 1 Classification of research issues in community question answering

literature that has employed ML and DL. To achieve this goal, a systematic literature review was conducted. Specifically, the review seeks to synthesise CQA research's major themes related to the question module, the answer module, and the user module.

Such an effort is timely because systematic reviews on CQAs are far and few. As shown in Table 1, among the handful of works that exist, most are quite dated [10–14] and hence capture the literature at a point in time when CQAs were not as matured and well-established as they currently are [15]. Some of the more recent literature surveys had a myopic focus. For example, some were solely focussed on the use of NLP to generate automatic answers [16, 17], which are different from user-generated answers. The one by Yuan et al. [25] focussed only on the issue of finding experts on CQAs. The survey [26] explored predictions of cyber-attacks utilising real-time twitter tracing recognition, which can be influenced. Similarly, motivating people to use information technology [27], Hash Hirschberg protein alignment utilising hyper threading [28] as well as personal privacy [29] and IoT usage evaluation [30] are all classical data dependent persuasive estimations. Yet, from 2005, a huge volume of articles have been published on CQAs, as shown in Figure 2. Some of the studies published post-2015 are informed by the latest progress in techniques such as DL [31–33]. These have often used transformer-based models such as bidirectional encoder representations from transformers (BERT), robustly optimised BERT (RoBERT), multimodal fusion transformer with BERT (MMFT-BERT) and other variants [34–38]. The BERT model is not limited to textual data only but is also used in visual question-answering systems [37]. Transformer-based models have been shown to speed up CQA systems [38]. For these reasons, now seems to be an opportune moment to carry out a systematic review on CQAs, encompassing not only ML-based but also DL-based approaches.

The main purpose of this survey paper is to provide a summary of existing research with an overview of the state-of-the-art literature and related theories in the domain of CQA with recent computational technologies. The survey can serve as a comprehensive guide for scholars who are interested in investigating issues related to CQAs. Our main contributions are as follows:

- We explored issues related to CQAs and categorised them into three modules: (i) the question module, (ii) the answer module, and (iii) the user module.
- We collected and synthesised the extant research on CQAs from a wide range of sources.

TABLE 1 Existing survey for community question-answering sites

Source	Theme	Systematic review	Models covered	Type of data
Kolomiyets et al. [10]	Information retrieval	Yes	Machine learning	CQA
Allam et al. [11]	Approaches used for community question-answering system	Yes	Machine learning	CQA
Gupta et al. [12]	Question-answering techniques	Yes	Machine learning	CQA
Dwivedi et al. [13]	Research and review of question-answering system	No	Machine learning	CQA
Bouziane et al. [14]	Research and trends in question-answering system	No	Machine learning	CQA
Lie et al. [16]	Methods and trend used in community question answering	No	Machine learning	CQA
Srba et al. [17]	Classification of approaches for community question answering	Yes	Machine learning	CQA
Wu et al. [18]	Visual question answering	Yes	Machine learning	VQA
Diefenbach et al. [19]	Core techniques used for community question answering	Yes	Machine learning	CQA
Soares et al. [20]	Existing approaches for community question answering	Yes	Machine learning	CQA
Mishra et al. [21]	Overview of community question-answering (CQA) websites	Yes	Machine learning	CQA
Ahmad et al. [22]	Review on question-answer	Yes	Machine learning	CQA
Yang et al. [23]	Expert recommendation	Yes	Machine learning and deep learning	CQA
Khusro et al. [24]	Overview of social question-answering websites	Yes	Analytical	CQA
Yuan et al. [25]	Finding experts		Machine learning and deep learning	CQA
Our work	Studies of community question answering	Yes	Machine learning and deep learning	CQA

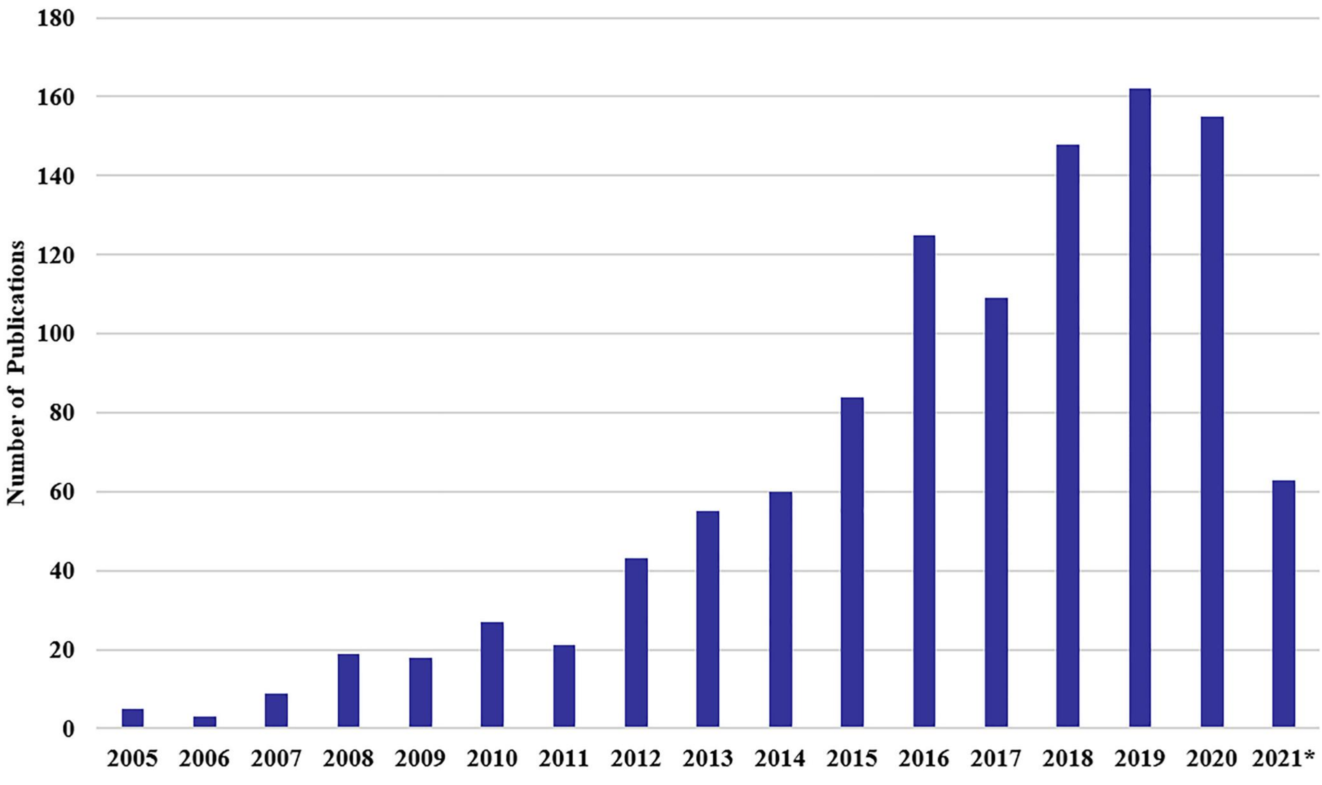

FIGURE 2 Number of research articles published over the year to resolve the different issues of the community question-answering system

- The key studies on each of the three modules are addressed separately with their findings and limitations.
- The challenges and open research issues discussed in this paper will enable future researchers to further enrich the CQA literature in the future.

The rest of the paper is organised as follows. Section 2 presents the process of collecting data. Section 2.1 introduces DL models, and Section 2.2 presents the categorisation of the collected data. In Section 3, issues related to the question module are discussed. Section 4 discusses the issues of the answer module whereas Section 5 describes the user module issues. Section 6 presents the research agenda for the different modules. Finally, Section 7 offers a conclusion.

2 | METHODOLOGY

The major part of the current work was to gather the relevant research articles from various sources and classify them into the categories of questions, answers or users—depending on their thematic focus. After the categorisation, we discuss the problems they have addressed, their methodology, outcomes, shortcomings, and future research. We relied on multiple databases such as ACM Digital Library, IEEE Xplore, SpringerLink, ScienceDirect, and Scopus. These were searched using the following commonly used keywords in the CQA literature: 'CQA', 'question answering', 'social question answering', 'expert users', 'question quality', 'answer ranking', 'reputation collectors, 'stack overflow', 'stack exchange', 'Yahoo! answers' and 'question-answering'. Moreover, we manually selected only those articles that were specifically devoted to studying CQAs using ML and DL from the search results. In addition, to get more coverage, we manually investigated the references of the selected articles. This data collection approach ensures greater comprehensiveness of our literature search compared with traditional information gathering [39] and data dissemination [40].

A total of 1106 articles were collected from the above-mentioned sources. However, during the initial screening of the downloaded articles, it was noticed that not all articles were relevant to our work. Hence, irrelevant articles are excluded based on the following criteria:

(i) Written in non-English language,
(ii) Studied non-textual data such as images and video,
(iii) Published in conferences,
(iv) Not specifically relevant to issues of questions/answers/users, and
(v) Had no proper implementation using ML and/or DL.

Then, the remaining articles were categorised into the three modules, that is, questions, answers, and users, with the help of three independent researchers. All conflicts were resolved through discussion until a unanimous agreement was achieved. After data cleansing, 133 articles were included in the final review. The steps followed to identify articles are shown in Figure 3.

2.1 | DL models and evaluation metrics

We review the CQA literature that has used either ML or DL. For brevity, we refrain from explaining the ML models in this paper. However, the interested readers may refer to Kowsari

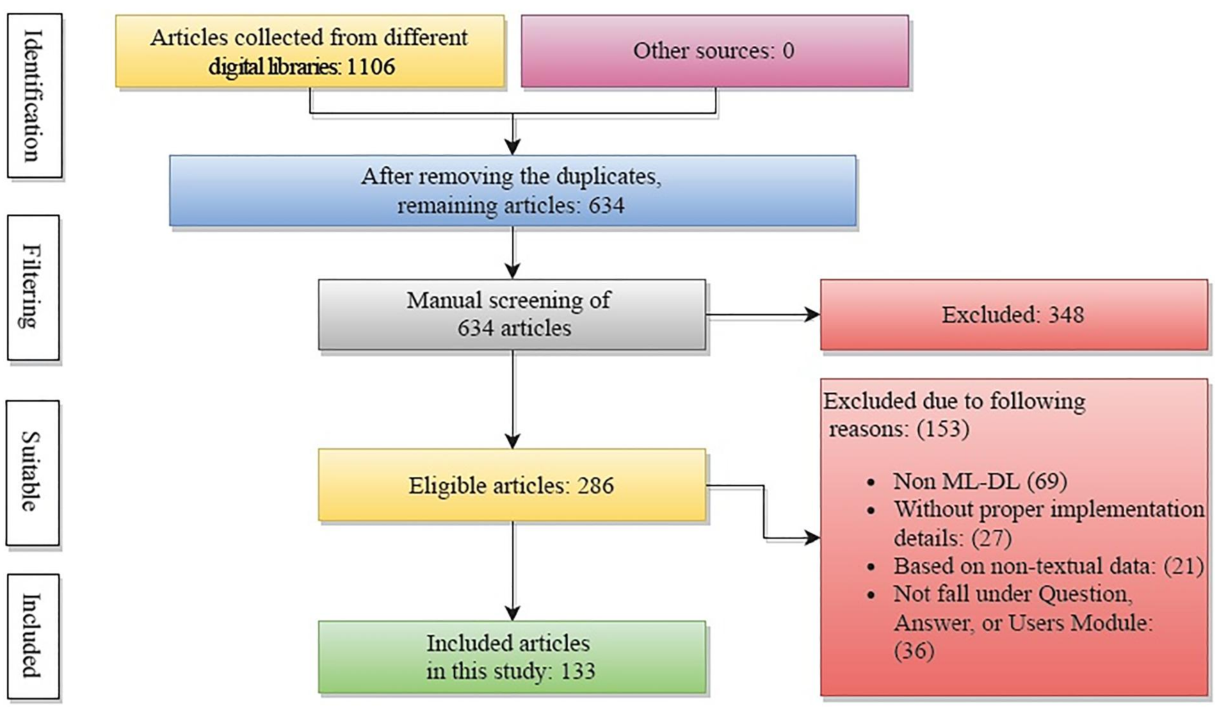

FIGURE 3 Article identification process

et al. [41]. Nonetheless, DL models such as convolutional neural network (CNN) and long-short term memory (LSTM) are more contemporary. These are therefore briefly explained below to familiarise readers.

DL-based frameworks have become extremely popular in recent years primarily because they remove the dependency on hand-crafted features. Manually extracted features could be biased or vary based on researchers' expertise. It is also tedious and time consuming to find the features that help achieve maximal performance. Moreover, hand-crafted features are problem dependent. For every problem, a separate feature engineering exercise is needed to extract the problem-based features.

In contrast, DL models such as CNN [42], RNN, LSTM [43], and bidirectional LSTM [44] models are able to extract the concealed patterns present in the text by utilising their multiple intermediate layers. Since they feed on textual data as input, the developed models can be reused for other similar tasks easily. In addition, with unstructured data, recent research has shown DL to outperform ML [33].

However, deep neural networks have some limitations, including the requirements of large amounts of samples for training, a high configuration-based system for training models, and more time to build the model. On the other hand, traditional ML algorithms are fast to train and can be used with a low amount of training samples. They show fast convergence and can be implemented on a machine with an average configuration. Besides, to use traditional ML classifiers, labelled data is required. But labelled data is not always available. DL models win the race here also as they perform well with unlabelled data. While both ML and DL have pros and cons, DL is increasingly preferred in the CQA literature in recent years.

In the subsequent subsections, the working principle of DL-based models such as CNN and LSTM with word embedding techniques will be discussed. CNN and LSTM models are most frequently used in the CQA literature. The CNN model helps capture the overall semantics of sentences, while the LSTM model is best suited to time series data. Other deep neural network frameworks, such as RNN, Bi-LSTM, LSTM-based autoencoder, and so on, have also been used but are less frequent [41].

2.1.1 | Convolutional neural networks

CNN is widely used to solve the problem related to images and text [42]. The complete working principle of the CNN model for text classification is shown in Figure 4. The input layer of the CNN network receives text data as input, and with the help of word embedding techniques, the input layer creates a matrix for each question and answer. Further, with n-gram filters, the convolution operation is performed over the sentence matrix to extract the unique and hidden word features from the textual contents [43]. To reduce the computation overhead of the network, the pooling layer pooled out the important features from the list from a fixed window size. At last, the pooled features are flattened and fed to the dense layer to perform the prediction task [44].

2.1.2 | Long-short term memory

The LSTM model is capable of remembering long sequences and hence is particularly suited to solve time series problems such as stock market price prediction, weather forecasting, and others. In recent years, it has also been used to address question-answering problems [33]. The LSTM model mainly consists of four components namely (i) forget gate (ft), (ii) input gate (it), (iii) cell state (Ct) and (iv) output gate (Ot) [43].

The cell state keeps the processed information; this is also known as the memory of the network. The forget gate is responsible for removing the unimportant message from the memory and passing the relevant one to the next state. To update the cell state information, an input gate is used. The content of the hidden state is decided by the output gate. Mathematically these components are defined below:

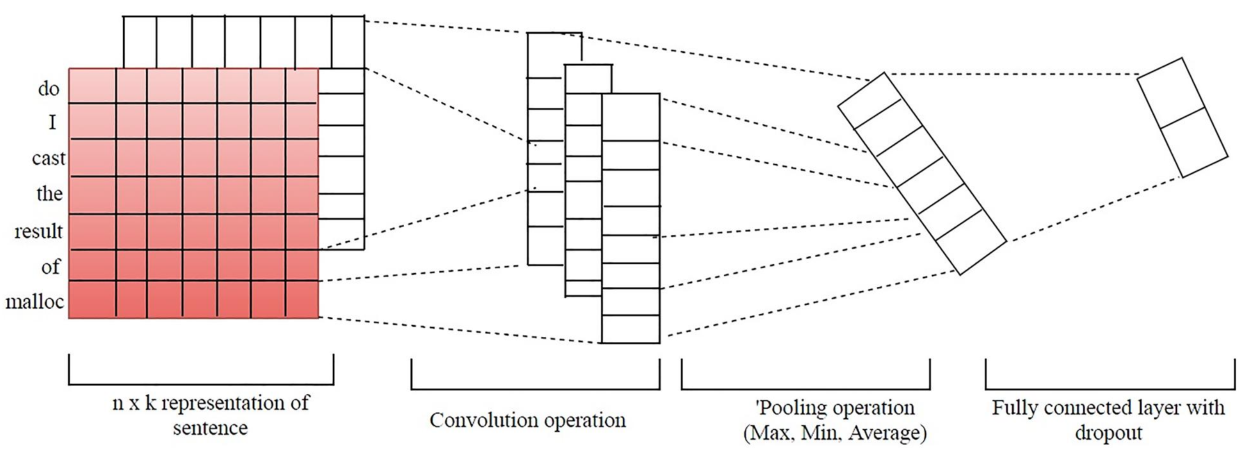

FIGURE 4 Convolutional neural network model for text classification [45]

– Input gate *(it)*: $it = \sigma(Wi[ht-1, lt] + \beta i)$
– Output gate *(Ot)*: $ot = \sigma(Wo[ht-1, lt] + \beta o)$
– Forget gate *(ft)*:$ft = \sigma(Wf\,[pt-1, lt] + \beta f)$
– Cell state *(Ct)*: $ct = tanh(wc[ht-1, lt] + \beta c)$

Here, Wi, Wo, Wf, and Wc are the weight matrices and the biases of the neurons are βi, βo, βf, *and* βc.

During the training process, the relevant information from the word sequences are kept by these gates and further used for prediction. A unit of LSTM network can be seen from Figure 5, where $Ct-1$ and Ct represent the cell state. The $ht-1$ and ht represent the hidden layer outputs at $t-1$ and tth time stamp, respectively. The Xt is the input word at time t. The calculated cell state Ct and hidden state ht values are fetched to the next unit of LSTM network as input.

2.1.3 | Word embedding techniques

DL models require a word vector as input. To convert the data into the required format, word embedding techniques are used [33]. Word embedding is a type of word representation that allows words with similar meanings to have a similar representation. The widely used word embedding techniques include (i) Skip Gram [47] and (ii) Continuous Bag of Word (CBOW) model [47]. Researchers also use the pre-trained word vectors such as:

(i) GloVe [48], which was created using Wikipedia data. The GloVe is available with different dimensions such as 50, 100, 200 and 300, meaning every word mapped with 50, 100, 200 and 300 other similar words, respectively,
(ii) GloVe on Twitter: This word embedding was created using Twitter dataset. Research on Twitter uses this embedding, whereas, for other datasets, Wikipedia-based word embedding is preferred, as reported online: https://nlp.stanford.edu/projects/glove/

2.1.4 | Evaluation metrics

To evaluate the performance of the proposed models, a number of different metrics are commonly used. The most commonly used metrics that are used for binary classification, multi-class classification and for ranking problems in CQA systems are listed below. Considering $i \in$ *number of classes* [49, 50].

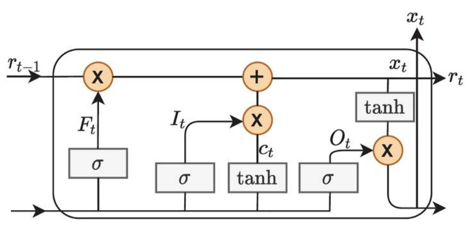

FIGURE 5 A long-short term memory unit with different components [46]

– *Precision*: The fraction of truly predicted data to the number of data of the retrieved among the total actual data.

$$\text{Precision} = \text{True positive}/(\text{True positive} + \text{False positive}) \tag{1}$$

– *Recall*: It is defined as the truly predicted data to the total number of actual data instances

$$\text{Recall} = \text{True positive}/(\text{True positive} + \text{False negative}) \tag{2}$$

– *F1-Score*: The harmonic mean of Precision Equation (1) and Recall Equation (2) is defined as F1-Score.

$$F1 - \text{Score} = 2 \times \text{Precision} \times \text{Recall}/(\text{Precision} + \text{Recall}) \tag{3}$$

– *Micro Precision*: it is the fraction of predicted true positive instances of all the classes to the retrieved instances of true positive and false positive.

$$\text{micro Precision } (P\mu) = \frac{\sum_i (TP)_i}{\sum_i ((TP)_i + (FP)_i)} \tag{4}$$

– *Micro Recall*: it is the fraction of predicted true positive instances of all the classes to the retrieved instances of true positive and false negative.

$$\text{micro Recall } (R\mu) = \frac{\sum_i (TP)_i}{\sum_i ((TP)_i + (FN)_i)} \tag{5}$$

– *Micro F1-score* (Fμ-score) is computed as: $F\mu - score = 2 \times P\mu \times R\mu/(P\mu + R\mu)$

– Micro receiver operating characteristic (ROC) curve: The plot between micro true positive rate and micro false positive rate, where true positive is projected on Y-axis and false positive projected on X-axis.

$$\text{micro TPR} = \frac{\sum_i (TP)_i}{\sum_i ((TP)_i + (FN)_i)} \tag{6}$$

$$\text{micro FPR} = \frac{\sum_i (FP)_i}{\sum_i ((FP)_i + (TN)_i)} \tag{7}$$

– Hamming Score: It is calculated by the fraction of wrong labels to the total number of labels as Equation (8), where yi,j is the target and zi,j is the prediction.

$$\frac{1}{|N| \times |L|} \sum_{i=1}^{|N|} \sum_{j=1}^{|L|} XOR\,(y_{i,j}, z_{i,j}) \tag{8}$$

2.2 | Categorisation of the literature

As shown in Table 2, the collected articles are categorised into three main modules: (i) the question module, (ii) the answer module, and (iii) the user module. Furthermore, the question module was classified into duplicate questions, question answerability, question routing, and question quality. The answer module was classified into answer quality, best answer prediction, and answer ranking. The user module was classified into two main categories: finding expert users and reputation collectors. In Table 2, we provide a detailed criterion about classification.

3 | QUESTION-RELATED STUDIES ON CQAs

CQAs invite people to ask questions in natural language. Most platforms such as SE provide clear instructions to the questioner about how to post good quality and non-repetitive questions. Yet, a huge number of duplicates, off-topic, and non-constructive questions are posted. The site moderators regularly monitor such questions and mark them closed from time to time. However, it is a daunting task for humans [72]. The research community has been trying to mitigate these issues on duplicate questions detection [73, 74], question unanswerability [75–77], question routing [78], and low-quality question detection [79].

3.1 | Duplicate question detection

CQAs often repeat questions that are repetitive. This has given rise to a body of research on duplicate question detection [72]. Sometimes, the question text of the duplicate question may have a good similarity with the previous question, but other times, it may contain different words also. Hence, identifying duplicate questions is a challenging task. Zhang et al. [73] found that the duplicate question pair contain similar tags and similar latent meaning irrespective of differences in words. They used these observations as a key factor to detect duplicate questions on SO. Four different similarity measures such as (i) title similarity, (ii) description similarity, (iii) topic similarity, and (iv) tag similarity were used to achieve a recall value of 0.638. Ahasanuzzaman et al. [74] proposed a classification and ranking model called Dupe. Their model works in three phases. In phase one, features were extracted. Phase two was to classify the question into two classes, and finally, in the third phase, the duplicate questions were ranked based on their probability value. They achieved a recall value of 0.6611 for the best case.

TABLE 2 Classification of community question-answering (CQA) issues

Module	Classification	Explanation	Source
Question	Duplicate questions	On CQAs, people are free to post any number of questions, and hence many questions are posted repetitively. The duplicate questions usually convey a similar meaning but with different words.	[49, 50]
	Question answerability	Users often post questions that cannot be answered. In other words, they lack answerability. This could be because they are unclear or simply seeking subjective opinion.	[51, 52],
	Question routing	To get high-quality answers to questions, CQAs route the questions to expert users. The routing of the question depends on several parameters, including tag of the question and the question's domain.	[53, 54]
	Question quality	CQAs are not restricted to their users about what to post. Hence, many users post questions that are inappropriate and subpar.	[55–57]
Answer	Answer quality	Each question on CQAs can receive multiple answers. The quality of these answers may vary from low to high in terms of content representation and explanation. Low-quality answers should not be allowed to be posted on a question that already has sufficient high-quality answers.	[58, 59]
	Best answer	Many times, questions on CQAs receive multiple answers from peers. The automated model helps to decide among the listed answers which should be the best one.	[60–62]
	Answer ranking	If the answers receive on a question are more, then all answers cannot be displayed on a single webpage. Hence, to serve the visitors, it is needed to list the answer by following their content. Currently, the vote-based mechanism opted to list the answer on popular CQAs	[63–65, 67, 68]
User	Expert users	To get high-quality answers for posted questions, expert users are needed. They can be identified through analysing patterns of user activities on CQAs.	[25, 57, 69]
	Reputation collectors	A handful of users on CQAs are involved in diluting the quality of content and receiving a high reputation in the early phase. Such users are termed as reputation collectors. They are mainly targeting low-quality questions to answer.	[5, 70, 71]

Zhang et al. [77] extended their model by including more textual and non-textual features. They found that in the programing domain, duplicate questions may contain code snippets along with the NLP text. Hence, Ref. [77] extracted many features from the newly posted question's title, body, and tags, including the code snippet and used similarity measure between the new post and existing question's title, body, and tags to find the duplicates. Zhang et al. [49] proposed another model called PCQADup. Their model generated (i) vector similarity, (ii) topical similarity, and (iii) association score. Vector similarity contains a question representation in a high dimensional continuous vector space. The LDA algorithm was used for topical similarity. For this, the question's title, body, and text data were used. The third feature was an association score; they mined association pairs and generated the features from them.

Chen et al. [78] proposed a question retrieval model to reuse the previously solved question-answer pairs. Silva et al. [79] claimed a 20% improvement in the accuracy in duplicate question detection. Two datasets are used for the research, (i) Ask Ubuntu and (ii) Meta SE dataset having 31 and 25k data, respectively. They achieved 92% average accuracy on both the dataset using a DL-based CNN model. Godbole et al. [80] used a combination of Siamese neural network and random forest classifier to detect the duplicate questions and achieved the loss value of 0.24.

Shah et al. [81] used a DL-based Bi-LSTM autoencoder model to detect the duplicates. SE and Quora datasets were used for the research study and it achieved better accuracy with the Bi-LSTM model as compared to the baseline models. Figueroa et al. [82] predicted the anchored temporal questions by combining heterogeneous sources. On CQAs, once a questioner confirms an answer as the best answer, the question-answer pair moves to the platform repository or archive. The archive is further used to find any similar posts in the future. The diverse output of the question helps the questioner to acquire knowledge in different areas. Relatively, Abric et al. [83] found the novice users are usually posting duplicate posts. However, the answers that the duplicate posts receive are different from the original post. Hence, it helps the users get additional information from the posted answers. An attention-based deep neural network framework was proposed by Liang et al. [76] for duplicate prediction on CQAs. Another reason behind closing a duplicate question is temporality. Some questions are temporal in nature and hence their answers too. If the question is outdated, their answer is also no longer valid. Hence it is needed to close them. For example, if someone was asked a question about a date of event in the year 2021, then the answer posted on this question may not be valid from the year 2022 onwards [84, 85]. The key research with additional finding from the existing research are shown in Table 3.

3.2 | Question un-answerability

CQAs attract several questions that are unlikely to receive answers. Unanswered questions are quite frustrating for questioners. To resolve the question un-answerability problem, Yang et al. [75] suggested a classification model in 2011. Their model predicted whether a question will receive the answer(s) or remain unanswered. They used the YA dataset, having a total of 76,251 questions; among them, 10,424 (13.67%) questions were unanswered. Features such as

TABLE 3 A summary of research studies on duplicate post detection

Source	Problem	Approach	Results	Dataset
Duplicate question detection				
[86]	Duplicate post detection	Classification	F1-score: 0.65	Twitter
[73]	Duplicate question detection in stack overflow	DupPredictor	Recall rate @20: 63.8%	Stack overflow
[74]	Duplicate questions	BM25, Dupe, DupePredictor and stack overflow	Recall-rate@20: 66.10%	Stack overflow
[49]	Duplicate post detection	Machine learning-based classifiers	Recall: 0.87, and F1-Score: 0.90	Stack overflow
[77]	Duplicate post detection	Classification	Recall: 0.96	Stack overflow
[87]	Duplicate question detection	Classification	F1-score: 0.68	Stack exchange
[88]	Finding near-duplicate post	Clustering and classification	Precision: 0.74	Software developer network
[89]	Duplicate programing question detection	Multi-layer perceptron	Recall: 0.89 and F1-Score: 0.90	Stack overflow
[90]	Prediction of linked questions	Deep learning-based classification model	Accuracy: 84.10%	Stack overflow
[91]	Prediction of related questions	Deep learning-based classification model	Precision: 0.89, recall: 0.89, and F1-Score: 0.89	Stack overflow
[33]	Prediction of closed questions	Deep learning-based multi-class classification model	Precision: 0.47, recall: 0.48, and F1-Score: 0.48	Stack overflow

'question length, category matching, asker history, question-posting time, question subjectivity, and polite words' extracted from the dataset and fed into four different classifiers, namely (i) Naive Bayes, (ii) Decision tree, (iii) AdaBoost, and (iv) SVM; among all the SVM classifier yielded the best performance. The work proposed by Yang et al. [75] was extended by Dror et al. [92] by predicting the answer counts a question may receive. A questioner posted the question and expected to receive the peers' answers; however, sometimes questions not received any answers for a long time. They said it's better to inform the questioner how many answers may be received from the peers instead of simply saying get answers. To do this, they extracted the features from question metadata, question content, and user data and designed a regression model. Their model achieved the root mean square error (RMSE) value of 4.606 and the Pearson correlation 0.620 for the best case. Ref. [51] found that the number of unanswered questions has increased significantly in recent years on SO. The questions are marked as closed by the moderator if it falls under the following categories: (i) duplicate to an earlier question, (ii) remained unanswered for a long time (iii) off-topic, (iv) not constructive, and (v) not a real question. Based on our knowledge, researchers address two issues among the five, such as duplicate questions and question un-answerability and the other three issues remained untouched, as listed in Table 4. In the future, models can be developed to address the other close-ability issues. In addition, upcoming relative research can step into physical interaction [52] and asymmetric agent connections [53] as well as microscopic large-scale modelling [54] aiming to find attractive remarks.

3.3 | Question routing

Question routing is an essential part of any CQAs. It plays a major role in providing a better answer to the community's question. The question routing technique's main principle is to identify the potential answerer for the newly posted questions. Many research studies have been reported to resolve this issue [25, 96, 97]. User posted the question and wait for the answer till the right users have seen their question. This mechanism leads to let answer or sometimes no answer to the posted question. To overcome this issue, Zhou et al. [97] developed a model using the forum's content and structure for pushing the newly posted question to the right users. The routing mechanism suggested by Li et al. [96] consider multiple factors, including the quality of answers posted by the users, number of answers submitted, tags and others, to predict the expertise. They confirmed that the routed question receives at least one answer if it is recommended to the top 20 predicted experts. The primary task of the computational model is to understand the query raised by the questioner. Often, the asker fails to frame the question properly and finally leads to either remaining unanswered or receiving few answers after a long wait. A model was suggested in Ref. [98] to re-write the ambiguous questions to correctly interpret the context.

Zhou et al. [99] proposed a classification using different set of features extracted from the question, such as title length, length of question body, and type of questions (what, why, who, how, where and when). From the profile, they extracted the features as the member since, total reputation points, number of posted questions, answers, the best answers. The features from question-user relations include the user is a top

TABLE 4 A summary of research studies on question un-answerability and low-quality question detection

Source	Problem definition	Approach	Results	Dataset
Question un-answerability				
[75]	Analysing and predicting not answered questions	Naive Bayes, decision tree, Adaboost and SVM classifier	F1-score: 0.32	Yahoo! answers
[92]	Predicting question answerability	Regression model	RMSE: 4.61 and Pearson correlation: 0.62	Yahoo! answers
[51]	Finding unanswered questions on stack overflow	Random forest, J48 classifier	Precision: 0.38 and recall: 0.45	Stack overflow
Low-quality post detection				
[93]	Analysing and predicting closed questions	Classification	Accuracy: 70.3%	Stack overflow
[59]	Improving low-quality post detection on stack overflow	Genetic model	Precision: 0.68	Stack overflow
[58]	Predicting deleted questions	Classification	Accuracy: 66%	Stack overflow
[60]	Why is stack overflow failing?	Analytical	Users such as noobs, help vampire, reputation collectors are the main source of low-quality content on SO.	Stack overflow
[94]	Deleted question prediction	Delpredictor	Precision: 0.52, Recall: 0.66, and F1-Score: 0.54	Stack overflow
[95]	Question quality analysis and prediction	CSMRLP	Precision: 0.84, recall: 0.85, and F1-Score: 0.84	Yahoo! answers

contributor or not to the question domain. Using these features, they achieved 76.89% of accuracy for the best case. Chang et al. [100] identified the potential collaborators for the questions by considering the user's availability, compatibility, and expertise for the recommendation. The experimental results on the SO dataset confirmed that the proposed collaborative approach yielded better performance. By utilising the non-QA data of the users, Srba et al. [101] proposed a model for question routing. They analysed the user data from CQAs as well as other platforms such as microblogs, blogs, and social networking sites. These data were used to estimate the user's interest and their expertise domain early. This will help to route the question to more experts.

Liang et al. [102] investigated the social role, interactions, and sustainability of community users on CQAs. The number of unanswered questions negatively impacts on future incoming questions. As a result, it may be possible that users did not post anymore on the particular websites if they found many of the earlier posted questions have remained unanswered. The questions and answers are the important resources of any CQAs. Sustainable CQAs maintain the rate of users' questions and answers. Before allocating a question to the users' their activeness needs to be checked, which improves sustainability. Li et al. [103] proposed NetRank model for question routing. Earlier models rank the users based on the question content, and hence the questioner information is ignored.

Yueliu et al. [104] extended the work of Ref. [56] and proposed a model that predicted the response time of the posted question. They said that earlier models are unable to give the guaranteed time frame to the questioner about the answerability to their question. However, their model helps the questioner to get a response from the expert users. Wasim et al. [105] proposed a model for the biomedical domain. The biomedical question categorisation is challenging as a question may belong to more than one class. The author developed a corpus for this task and using ML-based classifiers, achieved good classification accuracy. Visual QA models are popular in medical domain [106] but consist of many challenges. A model developed by Gupta et al. [107] address the issues of visual medical question-answering system. The end-users are expecting a quick answer to their queries from medical assistance. However, the models need to understand the queries to provide quick answers. Sometimes users' queries can be answered in one word, like yes or no, but for some questions, many need more information to provide the answer. Authors developed the model to handle the challenges of identifying types of the question in their research. A hybrid intelligent model was proposed by Belhadi et al. [108] for medical QA systems. Their model was capable of finding the relevant information from the medical data, recognise the patient activity, and retrieve the most relevant images from the medical database, different than traditional securing medical images [109] and cybercrime protections technologies [110] as well as stego-crypto protection schemes [111].

There were many pieces of research that have been introduced in the last decade for question routing. Still, it is not fulfilling the current needs. Many questions remained unanswered or received a very late answer. Hence, the CQA platform needed a large number of experts who can provide quick answers. To achieve this goal, instead of using the CQAs data only, the external resources of the registered users can be used [101]. This will help to fulfil the current need for experts for the CQA system.

3.4 | Question quality

Nowadays, CQAs are experiencing a surge in subpar content. The low-quality content on the SO website increased to 16.83% in 2016 compared to 4.11% in 2011 [60]. Espina et al. [112] focussed on why a question was asked on CQAs. Ponzanelli et al. [59] used the SO dataset to develope the low-quality post-detection model. They extracted 40 different features, which belong to (i) contextual features such as title length, body length, title body similarity, tags count, Flesch reading ease score, Gunning fog index, and others and (ii) Community-related features including closed votes, upvotes, downvotes, reopen votes, total badges, answer badge count, and others. With these feature set, they predict the low-quality post with 85.15% of precision value for the best case.

Correa and Sureka [58] found eight percent low-quality questions posted on SO. A question was deleted based on the negative votes given by the peer users. However, their analysis reveals that 75% of the questions were never voted for deletion. Hence, most of the questions were manually evaluated and deleted by the moderators only. A model called DelPredictor was proposed by Xia et al. [94], which predicted that the newly posted question would be deleted or not. DelPridictor was the combination of text processing and classification techniques. Text processing includes the extraction of textual features, then a classifier was used to train the model using the textual features, and finally, a composite classifier was used for predictions. DelPredictor achieved the F1-score of 0.54, which outperformed the model developed by Kim [45].

The questions, which are either duplicate, remain unanswered, or duplicate to previously posted questions, were closed by the site moderators. After analysing the post of multiple community sites of SE such as Programmer, Mathematics, Ask Ubuntu, and others, the highest number of the closed questions were found on topic Programmer as 25.64%, and the lowest number was on topic Code Review as 2.28% [33]. These statistics indicate that the CQA is overloaded with closed questions, which affect the site content quality as well as increase the moderator overhead [93].

The off-topic and other low-quality questions are marked closed by the site moderator. A model developed by Correa and Sureka [93] helps the moderator by predicting the closeability of the question at the beginning of the post. The best results reported by them with stochastic gradient boosting tree (SGBT) classifier was 70.3% in terms of accuracy measure. Another model was proposed by Roy and Singh [33] to predict the closed questions with their possible reasons. They used a dataset having five classes in which a class represents the open question category, and the remaining four classes belonged to a reason for closed questions. They used CNN and LSTM models for their work and achieved the best F1-Score as 0.48. Another model was developed using the coupled semi-supervised mutual reinforcement-based label propagation to predict the question quality on YA [95]. They consider the correlation between the question quality and answer quality and find the probability of obtaining a high-quality answer. They extracted several features from questions, askers, categories, and answers. A semi-supervised model was suggested in Ref. [113] using sentiment analysis and dependency parsing as well as named entity recognition to find the relevant questions and answers from archives. The linguistic features extracted using morphological analysis helped to achieve better performance. Harper et al. [114] suggested a framework for identifying the informational and conversational questions category on CQA with a smaller set of data and it was extended by Guy et al. [115]. The authors have suggested a model with traditional machine learning models such as random forest, support vector machine, logistic regression, and others with the comparatively larger dataset used by Harper et al. [114]. They also suggested an ensemble framework to identify the different categorical questions on CQA such as informational, conversational and others.

As summary of this section, it discusses the research done using various computational models for solving the issues of question module including the duplication question (Section 3.1), question answerability (Section 3.2), low-quality post (Section 3.4), and question routing (Section 3.3). The researchers used various machine and DL-based models to address these issues. Still, a huge number of questions remained unanswered on CQA platforms. Therefore, we need a more robust computational model to find the answers and answerer, who provide the answer to the questions. The duplicate questions of different CQAs, such as Quora, SE, and SO are also an open issue [74]. Researchers said that one of the reasons behind the high volume of unanswered questions is the insufficient number of expert's availability [4]. Therefore, a model that identifies the active experts or predicts the experts from the existing users of the CQAs may reduce the number of unanswered questions. Hence, future researchers may choose any of these areas and provide a general model that works with any question-answering websites without any feature dependency.

4 | ANSWER-RELATED STUDIES ON CQAs

Every CQA platform has the job of providing high-quality answers to the question asked. The website moderators regularly make considerable efforts to ensure that every question receives high-quality answers from community users, but not always high-quality answers are received. In fact, the quality of the users' answers is very diverse. CQAs attract not only high-quality answers but also irrelevant contents such as spam and abuses [116]. Improper answers are the answers that do not apply to the question: either it is incomplete if the answer given does not solve the problem fully or incorrect if the information provided is no longer valid [117]. On CQAs, the number of these unsatisfactory answers is consistently growing, rendering the platforms ineffective. There are no standard guidelines given by either CQAs or the research community to evaluate the posted answer quality. Prior research has used various measures to define the relationship between question and answer content and evaluate the answer quality. Some of these measures include similarity, comprehensiveness, accuracy, questioner satisfaction [117], originality, relevance, and completeness [118]. According to Rodrigo et al. [119], CQAs help users by supplying the relevant answers to their questions. Still, at the moment, many of the answers provided by the CQAs are inaccurate, leading to a lack of user trust in the system. Consequently, they suggested a model, validating the posted answer. Their model helps reduce the number of incorrect answers on the CQAs. In the same way, there are several research findings on the quality of the answer. A detailed description list of key research on answer quality prediction is presented in Table 5.

4.1 | Answer quality

Answer quality assessment refers to the methods proposed to measure the content of individual answer quality (i.e. low-quality and high-quality). Researchers have introduced several methods to estimate the quality of the answers [119]. Jeon et al. [120] used statistical methods to evaluate answer quality. The non-textual features were then augmented with textual features such as answer's accuracy, completeness, and reasonableness of the answers by Blooma et al and Agichtein et al. [121, 122]. It was found that textual characteristics affected the quality of the good answer compared to non-textual features. Among the textual features, language and answer length were found less significant. The non-textual features such as the answerer's or the questioner's reputation and authority were found to not affect answer quality.

Zhu et al. [118] proposed a multidimensional regression model to evaluate the answer quality. They selected thirteen-dimensional features such as 'informativeness, politeness, completeness, readability, relevance, conciseness, truthfulness, level of detail, originality, objectivity, novelty, usefulness, and expertise'.

Toba et al. [123] proposed a hierarchical classifier for the prediction of answer quality. They found that a good-quality answer is typically a long answer that is completed with facts and written in a good format. Wu et al. [124] suggested an unsupervised model for predicting low-quality answers from CQAs. They found that many features from the answer text, such as length, type, similarity, n-grams are more prominent. Zhang et al. [125] has said that the social CQAs become a top option for users to exchange information. Using the support of 382 exchangers of information, they found that personalised suggestions, topical experience, and social interactivity support obtain quality content. Alternatively, this will help create a forum for high-quality knowledge-sharing.

TABLE 5 Summary of key research studies on answer quality prediction

Source	Problem definition	Approach	Results	Dataset
Answer quality				
[120]	Answer quality prediction	Entropy calculation	*p*-value: 0.007	Naver
[121]	Best answer prediction	Linear regression	R-square: 0.89	Yahoo! answers
[122]	Finding high-quality answers	Classification model	Precision: 0.97, recall: 0.92, and F1-Score: 0.94	Yahoo! answers
[118]	Answer quality prediction	Classification and regression models	Accuracy: 90%	Answerbag
[123]	Finding high-quality answers	Classification models	Accuracy: 79%	Yahoo! answers
[124]	Low-quality answer detection	Classification models	Accuracy: 86%	SemEval-2015 and YA
[127]	Answer quality prediction	LSTM network-based model	F1-score: 0.62 and accuracy: 75.20%	SemEval-2015, 2016
[126]	Answer quality prediction in health domain	Convolutional neural network	F1-score: 0.92 and accuracy: 93%	Self-prepared

Researchers also used the DL-based models [126] for answer quality prediction. Suggu et al. [127] used both hand-crafted and automatic features to predict the quality of the answers. They extracted a large number of hand-crafted features from the question-answer pair to concatenate with the features extracted using bi-directional LSTM models to predict the quality of the answer in the health domain. Hu et al. [126] proposed a multi-modal deep belief network model to evaluate the physician's answer in the health care domain. They introduced unique non-textual features such as 'keyword density, number of non-repeated words, number of a high-frequency medical term, sentiment polarity of the answer, and so on', and achieved the AUC value of 97.8% for the best case.

4.2 | Best answer prediction

On CQAs, many questions receive multiple answers; questioner has authenticity to choose the best answer among the received answers [128]. One of the models for best answer prediction was proposed by Lee et al. [129] using MSN question-answer (https://www.msn.com/en-in/) data with 1.3 Million answers. Their model was based on the voting system whereby every voter is assigned a weight. The voter who previously voted answers selected best answer receives a higher vote than others. The other papers viewed the best answer prediction as a classification problem. For example, to classify the answer as the best answer, Ref. [128] used the traditional ML model. For best answer prediction, they used features extracted by Zhu et al. [118] along with a few manually extracted features. The classifier results showed that the features taken manually were not adequate to find the best answers. To take the research further, Blooma et al. [130] used three different features: social, textual and content-appraisal. Their logistic regression analysis indicated that the content-based features are among the most important features for the given task. The content-based features were also used by Burel et al. [131] along with user and thread features. User feature includes user profile, age, posting rate, answer number, answer ratio, best answer ratio, and others. With the selected features and the decision tree classifier, the average F1-score of 0.83, 0.84 and 0.87 for SAP, the server fault, and cooking datasets are obtained, respectively.

Tin et al. [132] used the answers quality, the question-answer similarity, and the answer-answer similarity features to find the best answer. They used the SO dataset for their work and achieved 72.2% of accuracy for the best case.

They found that among all the contextual features the answer-answer similarity played the main role in predicting the best answer. Gkotsis et al. [63] said that similar accuracy (as of Ref. [131]) could be achieved only by using the textual features alone. Sakai et al. [117] claimed that Yahoo!'s best answer prediction system is fundamentally biased. The questioner himself chooses the correct answer; therefore, it relies solely on the questioner's experience. If the questioner has less experience, then a fairly low-quality answer may be chosen by users as the best answer. The same problem was also pointed out by Chen et al. [133], who said that the best answer was evaluated by expert users and found that 70% of the answers have less than standard quality. To find the best answers on CQA sites, Elalfy et al. [134] suggested a hybrid model. They used content features, question-answer, answer-answer features, and the user's reputation score for model development.

Zhou et al. [135] said that the best answer could be predicted more effectively by utilising user profile information features. Their analysis confirmed that the engagement-related features such as the point earned in a current week and the average points earned by the user per week also create a major role in predicting the quality of the answers. In addition, the answerer profile picture has a contribution to predicting this task. Molino et al. [136] used the users, network, and textual features to predict the best answer using Random Forest classifiers. They experimented with the YA dataset, having a total of 40 million instances. Their model outperforms the existing model by 26% of the precision rate. Bhatt et al. [137] suggested a model that fulfils the CQAs' semantic and lexical gaps. To accomplish this task, they implemented a phrase level and a token level attention strategy. Their model merged the features in vector space extracted from the questions, answers, and externally extracted features. Their model provides better performance on the WiKiQA dataset compared to the current system. Table 6 provides a briefing of the research studies on best answer prediction.

4.3 | Answer ranking

The answer ranking mechanism is used to rank the answers [138]. Two widely used answer-ranking approaches are (i) regression score and (ii) rank-learning. Hieber et al. [66] used question-answer similarity features to rank the answers. Li et al. [32] suggested knowledge map techniques to store and retrieve the best pair of questions and answers from the archive. Ref. [133] developed a model using the user's votes. Their analysis confirmed that the answer-ranking technique could be improved by removing the voting information. Dalip et al. [139] used a large number of features set from the question's answers and users' data to perform this task. Ginsca et al. [140] ranked the answers using features of the answers' profile information such as the age of the answerer, about me, and the link to the external platforms. The votes obtained by any post have been the primary measure to rank the answers.

Roy et al. [6] pointed out that the current voting-based answer-ranking system suffers from 'Matthew Effects' [141]. An answer posted earlier receives more votes as it is placed at the top and also gets more visitors. On the other hand, a newly posted good-quality answer will be added at the end of the list and hence does not get enough visitors to fetch vote. Authors suggested a 'promising tab,' in which the answer will be listed based on their predicted votes. Zhou et al. [31] used recurrent convolutional neural network (RCNN) to rank the answers. The question and their answers were combined to create question-answer pairs, which are used as input to the CNN model to capture the semantic features and then fed into the RNN model to rank them. Zhao et al. [142] suggested a model called heterogeneous asymmetric multi-faceted network learning which was based on the LSTM network. The earlier studies considered this problem as a matching task that uses the deep semantic matching model [143] to find the score of semantic relevance for ranking. They found a model achieving good accuracy only by considering the semantic relevancy. Their analysis also confirmed that the expert users' answers have more relevant than other users [144]. Chen et al. [145] proposed a positional attention-based RNN model for ranking the answers. They argued that if a question word occurred in an answer sentence, then the words near to that need more attention as they may contain more information. Based on this hypothesis, they proposed a positional-based RNN model incorporating the question word's positional context into the answer's attentive representation.

Lukovnikov et al. [146] proposed a GRU-based neural model to rank the answers of the simple factoid questions. They reported several challenges of the existing question-answering system including the lexical gap, ambiguity, and

TABLE 6 Summary of research studies on best answer prediction

Source	Problem definition	Approach	Results	Dataset
[118]	Answer quality prediction	Classification with regression	Accuracy: 90%	Answerbag
[154]	Predicting best answerers for new questions	LDA	Accuracy: 45%	Sina.com
[130]	Selection of best answer	Cohen kappa statistics measure	Cohen kappa: 0.86	Yahoo! answers
[131]	Automatic identification of the best answer	Classification model	Precision: 0.87, recall: 0.87, and F1-Score: 0.87	Stack exchange
[63]	Best answer prediction using linguistic features	Classification model	Precision: 0.84, recall: 0.70, and F1-Score: 0.76	Stack exchange
[117]	Graded-relevance matrix to find best answer	Statistical model	Hit ratio BA@hit: 80%–90%	Stack overflow
[133]	Automatic identification of best answer	Supervised model	MAP: 0.50–0.52	Stack overflow
[134]	A hybrid model to predict the best answer in health domain	Classification model	Precision: 0.88, recall: 0.88, and F1-Score: 0.88	Stack overflow
[126]	The best answer prediction in health domain	Deep belief network and convolutional neural network	Precision: 0.96, recall: 0.98, and F1-Score: 0.97	Stack overflow
[65]	The best answer prediction using heterogeneous data	Multi-view learning and classification	Precision: 0.52, recall: 0.67, and F1-Score: 0.59	Stack overflow
[64]	Predicting best answer in CQA	Machine learning and deep learning	MRR@50: 0.28 and Accuracy: 72%	Stack exchange

unknown knowledge boundaries. Nguyen et al. [147] used both the conventional and abstract features to rank the answers on the dataset of SemEval 2016 (http://alt.qcri.org/semeval2016/index.php?id=tasks). The conventional features include the ratio of the number of words between test questions and (i) their similar questions, (ii) related question's answers, a bag of words, noun overlap, word overlap, and others. To extract the abstract features from the text they used convolutional neural network [45] and Bi-LSTM [44] model. These features were concatenated together to rank the answers. Their model achieved an F1-Score of 78.43 and an MRR of 86.23 for the best case. Amancio et al. [148] suggested a model, which ranked the answer based on recency and quality of the content. The ranking models rely on a voting system that is biased, and hence, the suggested model is helpful to rank the answer on the CQA platform. Xie et al. [149] used an attention-based mechanism to rank the answers. They used SemEval 2016 and SemEval 2017 datasets for model training and testing purpose. Matsubara et al. [150] confirmed that transformer-based models such as BERT are capable of re-ranking of the answers. Wu et al. [151] suggested a model to label the answers posted on the CQA platform. The answers posted on non-factoid questions were ranked by Surdeanu et al. [152, 153].

As summary of this section, it discussed the issues of the answer module resolved with different computational approaches such as ML and DL. The CQAs estimate the answers' quality based on the number of votes they receive. This means that if an answer has got more votes, it is of high-quality compared to lesser voted answers. But an answer posted earlier has a greater chance of getting more votes. Therefore, even if the answer posted later has quality, it cannot get enough visitors because it is by default placed at last and thus receives fewer votes. A better model is needed to identify the factors of the high-quality answers concerning the question for proper ranking. Another issue is finding the best answer. A new model can be created to integrate a question text to assess the best answer, rather than looking at the number of votes received. Table 7 provides a briefing of the research studies on answer ranking.

5 | USER-RELATED STUDIES ON CQAs

The third module of the CQA site is the user module. Users of CQAs may be questioners, answerers, or commenters. Those who are able to provide high-quality answers are usually known as experts or topical experts [17]. Table 8 lists the summary of key research on finding the Expert Users. Jurczyk et al. [165] used the link analysis approach using the Hyperlink-Induced Topic Search (HITS) algorithm to identify the expert users of the CQAs. Zhang et al. [156] proposed an algorithm called 'ExpertiseRank', which was based on PageRank algorithm [166] to evaluate the expertise level of the users. With the help of two human evaluators, they manually rank the users according to their expertise. The expertise was decided based on their previous post history. For every user, the evaluator needs to check hundreds of answers and questions. Hence, they only ranked 135 users, which is less than the total number of users. Hence, it was cleared that the manual evaluation would not help to find expert users. Liu et al. [167] and Aslay et al. [157] proposed the system to find the expert users with the help of all available information. Their model finds the experts by giving them more priority (i) to the user who gave the best answer than that of the other answerer and (ii) to the answerer than the questioner. They found that their model significantly performs better than that of the earlier model on YA data [156, 165].

TABLE 7 A summary of key research studies on answer ranking

Source	Problem definition	Approach	Results	Dataset
Answer ranking prediction				
[139]	Predicting deleted questions	Classification	Accuracy: 66%	Stack overflow
[135]	Why is stack overflow failing?	Analytical	Users analysis of SO.	Stack overflow
[140]	User profiling for ranking answers	Ranking support vector machine	MRR: 0.46	Stack overflow
[66]	Improving ranking of answers	Support vector machine	MRR: 0.68	Stack overflow
[133]	Votes calibration in CQAs	Genetic model	Precision: 0.68	Stack overflow
[6]	Ranking high-quality answers	Classification and regression model	Precision: 0.72 recall: 0.90, F1-Score: 0.80, and RMSE: 1.92	Stack exchange
Answer ranking using deep learning approach				
[142]	Ranking answers	LSTM network-based model	Gain: 0.88, P1:0.56, and Accuracy: 49%	Quora
[145]	Ranking answers	Positional attention-based recurrent neural network	MAP: 0.78 and MRR: 0.85	TREC-QA
[146]	Ranking answers	Neural network- based GRU mode	-	Simple questions [155]
[147]	Ranking question-answer pairs	Bi-LSTM, CNN, and NLP	F1-score: 0.74 and MRR: 86.21	SemEval-2016

TABLE 8 A summary of key research on finding the expert users

Source	Problem definition	Approach	Results	Dataset
[156]	Expertise network in online communities	Expert ranking algorithm similar to HITS	Accuracy: 82%	Online java forum
[157]	Competition-based network for finding experts	Manual modelling	Accuracy: 54%	Yahoo! answers
[158]	Finding experts using profile tags	CTAR, LDA	Average precision (P@10): 72% and MRR: 0.46	Tianya wenda (Chinese QA site)
[144]	Identifying authoritative actors in QA	Page rank and Z-score	Average quality score: 0.77	Yahoo! answers
[159]	Early detection of expert users	Classification model	F-measure: 0.50	Turbo tax live community (TTLC)
[160]	Identifying potential experts	Classification model	F-measures: 0.72 (TTLC) and 0.91(SO)	TTLC and stack overflow
[161]	Expert finding on online community	Page rank	Correlation score: 0.90	Java online community
[162]	Context-based user ranking	Wordnet similarity	Score: 38%	AskMe forum
[163]	Hybrid expertise retrieval model	Multi-model approaches	Accuracy: 98%	Yahoo! answers
[164]	Expert finding using CNN	Convolutional neural network-based model	Success@n: 0.30	Stack overflow

5.1 | Topical expert identification

The previous approaches were focussed on identifying the users who have complete domain expertise rather than finding and ranking the topical experts. Topical experts have expertise on a particular topic only [4, 158, 168]. To rank the users based on their expertise in a particular category, Ref. [158] proposed a model called Category Relevancy-based Authority Ranking (CRAR). Their model works in two phases. First, it finds the relevant topics from the text using LDA topic model [168] and then, using the link analysis, they rank them. They used datasets collected from Yahoo! Answers (YA) and Tianya Wenda, a Chinese CQA, reported at http://answers.yahoo.com, and http://wenda.tianya.cn/, respectively. Their model achieved the MAP value of 0.80, 0.82, and MRR values of 0.95, 0.97 for YA and Tianya Wenda, respectively. Cai et al. [169] said that the earlier models used HITS, PageRank (link-based method) and Z-score (Equation 9) to evaluate the expertise.

$$Z - \text{score} = (na - nq) \Big/ \sqrt{(na + nq)} \qquad (9)$$

In Equation (9), na represents the number of answers posted, and nq is the number of questions posted. They also used the graph generated using the question-answer dataset and evaluated the answer's quality. The information about the user's answer quality is one of the critical information to compute the user's expertise. Experimental results on SO and Turbo Tax dataset, reported at https://ttlc.intuit.com/, show that their model works better than earlier models. Zhou et al. [170] said that the existing approaches used the link analysis techniques to identify the expert users in which the user's interest, expertise, and reputation were not considered. To overcome these limitations, they proposed a topic-sensitive probabilistic model, the extension of PageRank algorithm [166]. They used the YA dataset and found that their model outperformed several existing models. Tondulkar et al. [64] proposed a model to rank the expert users in CQAs. They used several models such as bag of words (BoW), TF-IDF, Word2Vec, Feature Sum, Tag only, and others to test the model accuracy in terms of MRR and found their model scores better as compared to the existing models by achieving the best MRR score (MRR@50) as 0.28.

Liu et al. [171] investigated the participation of active expert users of CQA websites. On stack overflow, they proposed a personalised activity level prediction model. They discovered that the two most suggestive indicators for the activity level prediction task are post-related features and positive response-related features. They suggested investigating the correlation between expert activity and their culture or demography. Ghasemi et al. [172] constructed a community graph in which they employed a node embedding strategy to locate related users. They also looked at lexical and semantic overlaps between the new question and existing expert responses. The proficiency of responses was used to assess an answerer's expertise. Kundu et al. [69] used data derived from the social behaviour network, user profiles, and question and answer language. Liu et al. [173] suggested a graph convolutional neural network-based model for expert prediction in the CQA website. They used the proposed graph to build a relationship between co-answerers if they had both answered the same question. The graph depicts how users' professional abilities are similar, which aids in expert discovery and recommendation.

Effect of user's demography, location, gender etc. on user's role in CQA: Fu et al. [174] looked into the evolution of the user's role in community question answering for the purpose of question recommendation. They presented a time-aware methodology for tracking the evolution of user roles.

Experiments using stack overflow data revealed that a time-aware user's role model improved question recommendation performance considerably. Saxena et al. [175] presented a user's social roles and their distribution on online question-answering platforms. They discovered that the relationship between different user roles, such as questioner, answerer, and viewer, and diverse user characteristics, such as gender, age, and country, is still unknown. Ford et al. [176] looked into women's engagement in stack overflow. They discovered that women who met other women were more inclined to participate in conversation sooner than women who did not. Figueroa et al. [177] looked into the automatic recognition of the asker's gender in community question-answering systems. They used ML algorithms to identify the asker's gender. Overall criteria including the asker's occupation, industry, and age as well as fine-grained categorisation of the question are important in correctly identifying the gender. Linguistic features of the question as well as the self-description of the asker were both useful. Finally, at the time of posting, Multinomial Bayes classifiers outperformed MaxEnt. Ford [178] proposed a gender difference on stack overflow. Wang [179] discovered that there is a reputation disparity between men and women based on stack overflow data, which is projected to expand in the future. As a result, women's involvement is threatened, and they are assigned to the low-reputation category, which can be the case studying the impact of corona disease on the feelings of twitter users during the hajj season [180]. According to Figueroa et al. [181], the user's demographic plays a vital part in tailoring content and boosting user experience, as can be explored in future research within real-life hotel services applicability [182]. Recently, Gor et al. [183] assessed the performance of a skewed QA dataset on a variety of applications based on gender, profession, and nationality. They discovered that, while datasets are imprecise and reflect social faults, there is no clear evidence that QA suffers from this bias for many demographic features. This demography analysis can be reported as real concern feeding artificial intelligence in proper exploration within [184]. Relatively, Figueroa et al. [185] suggested a multi-modal method for automatic age recognition across CQA platforms that uses text, image, and metadata [186].

Social influence on user's CQA contribution: In a social question-answer community, Dong et al. [187] explored the role of social influence on endorsement behaviour. To determine the relationship between social influence and social endorsement, they built a psychological choice model. They discovered that social endorsement is strongly connected with popularity and source credibility in social influence. Shi et al. [188] investigated whether previous contribution behaviours and community feedback influenced the features of users' subsequent CQA contributions, which did not put any attention to info secrecy or authenticity reported in Ref. [189]. Likewise, Li et al. [190] said that the private information disclosure on social media can be driven by voluntary sharing and mandatory disclosure.

5.2 | Reputation collectors and other inefficient users

While much research has shed light on expert user identification, Srba et al. [60] categorised inefficient CQA users in three different categories: (i) help vampires, (ii) noobs, and (iii) reputation collectors. Help vampires generally post questions without trying to find the answers from the existing database or other Internet sources. Their questions are often duplicated to an existing one. Their only intention is to get peers' answers but are not interested in sharing their knowledge by giving answers to other questions. Noobs post trivial or poor-quality questions on the site due to lack of expertise. As a result, CQAs are flooded with low-quality questions, making it difficult to identify interesting and informative questions. Reputation collectors target several low quality questions to answer to earn reputation and achieve the site's privilege.

According to Srba and Bielikova [60], the number of help vampires is usually constant, noobs increased from 6.23% to 10.11% from 2011 to 2014, and reputation collectors also increased from 4.11% to 5.98% from 2011 to 2014. Roy et al. [71] also investigated user behaviours on CQAs and found that reputation collectors mainly target either duplicate questions or low-quality questions to answer and gain reputation points. Reputation collectors are mainly involved in increasing the volume of low-quality content on CQAs [5]. Slag et al. [70] found many users posting their questions on CQAs and never coming back again. They termed such users as 1-day flies. Questions posted by these users were either removed by the community or by the site moderators to maintain the overall content quality.

6 | OPEN RESEARCH ISSUES, DIRECTIONS AND CHALLENGES

CQAs have been attracting huge attention from worldwide users. On these platforms, users ask a variety of questions with an expectation that someone will answer. Often, questioners post a question that may not have a valid answer; hence, it remains unanswered for a long time. Apart from this, the question may need domain expertise to answer. If a question does not fit in the domain of a particular platform, it is marked as closed. All such issues have existed from the beginning, and researchers have been suggesting many models to address them. However, due to the unstructured format of the data, it is a challenging task.

As shown in this review paper, all informed topics related to questions, answers, and users are still open for interested scholars in the future. Earlier, these CQA issues were addressed with manual feature engineering techniques and ML. However, DL-based models such as CNN, RNN, and LSTM networks are currently popular in the research community. These should be used more extensively going forward.

The descriptive statistics of the articles covered in the literature survey area is shown in Figure 6. As shown in

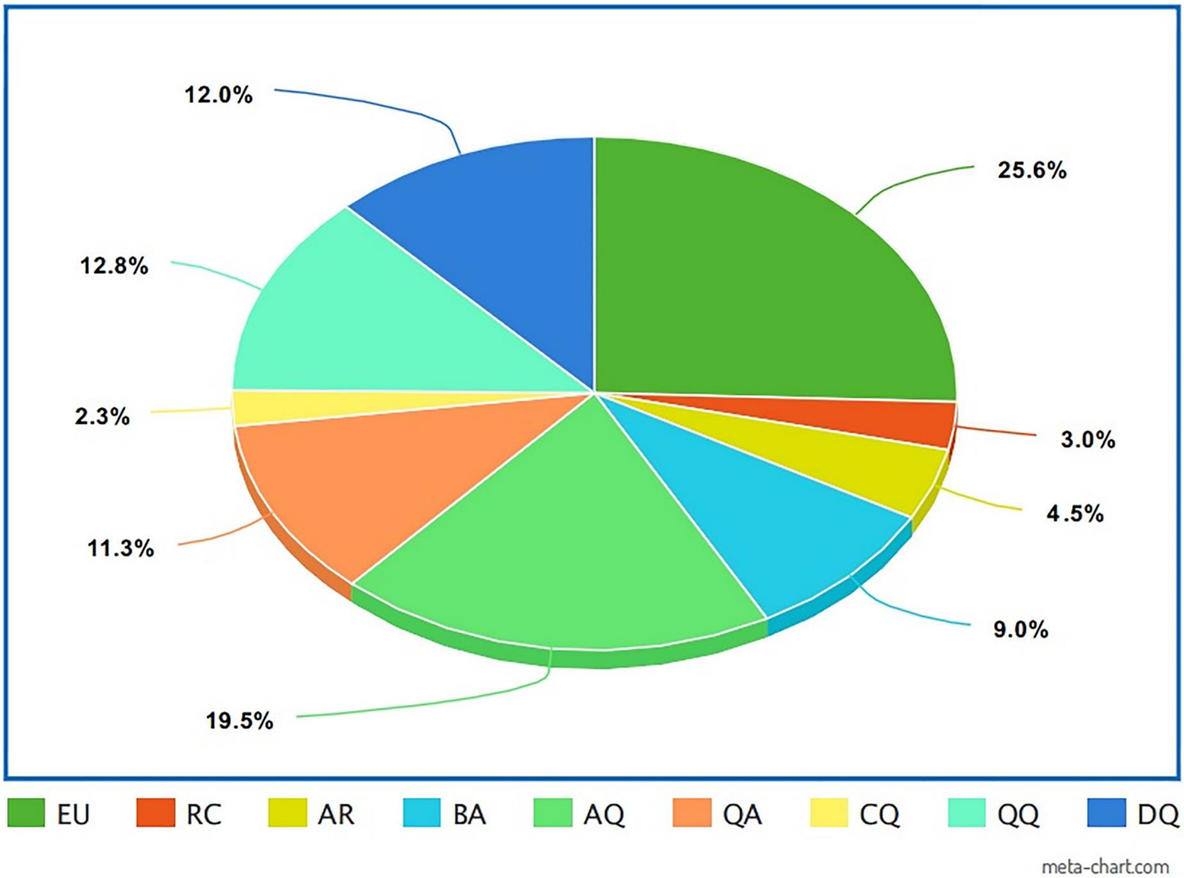

FIGURE 6 Number of publications published on different issues of community question answering. AQ, answer quality; AR, answer ranking; BA, best answer; CQ, closed question; DQ, duplicate question; EU, expert user; QA, question answerability; QQ, question quality; RC, reputation collectors

Figure 6, finding expert users (EU) has the highest number of articles. Finding answer quality (AQ) is the second most published topic. On the other hand, the closed question prediction (CQ) in the question module is the least published topic. However, it has been seen that in the last 5 years, the number of closed questions is increasing [33, 93]. Therefore, research needs to focus on this topic with a sense of urgency. The reputation collectors (RC) in the user module has the second-lowest number of articles. It has been recently found that malicious users can disrupt the seamless functioning of CQAs. Identifying such users and developing ways to stop them also need immediate attention, not as reported for increasing number of participants adopting counting-based secret sharing [191]. Besides topic modelling, the tag suggestion for a new question has attracted very limited research [192]. Hence, in the future, interested scholars may consider these topics to improve the content quality of CQAs. To continue the research on CQAs, the datasets listed in Table 9 can be considered by future researchers.

Apart from the above-mentioned issues, the challenges in CQAs research include the following:

(i) Question answering for E-commerce: Popular e-commerce platforms such as Flipkart and Amazon provide space for users' queries on the product webpage. If buyers have any question, they can post questions. These can be answered by either other customers or by the seller. The time gap between posting the questions and receiving the answers can be fairly long. This can be minimised by introducing an AI-based automated question-answering system. The main challenges for such models include the lack of a required dataset for model development.

(ii) Domain-specific interactive chat-bot system: This is another area that has not been explored. People are often completely dependent on search engines such as Google and Bing to get answers to their domain-specific queries as responses to such questions take a long time to come on CQAs. An interactive chat-bot system may be developed for domain-specific queries submitted on CQAs to reduce the turnaround time. The main challenge in developing such a system is the lack of training samples. This area needs more scholarly attention.

(iii) Language-independent CQAs: The most commonly studied CQAs use English language. Popular platforms could consider allowing questions and answers to be posted in any language for greater digital inclusivity. A major issue is the complexity of questions. Complex questions might not be successfully translated to other languages. Hence, it is important to develop language translation models for deployment on CQAs that will be

TABLE 9 List of popular datasets for question answering

Dataset	Source	Domain
Stack exchange	https://archive.org/download/stackexchange	Topical QA
SQuAD	https://rajpurkar.github.io/SQuAD-explorer/	Reading comprehension-based QA
Stack overflow	https://archive.org/details/documentation-dump.7z	Programing-based QA
Natural question	https://ai.google.com/research/NaturalQuestions	Open-domain QA
QuAC	https://quac.ai/	Dialogue-based QA
TWEETQA	https://tweetqa.github.io/	Social media-focussed QA
NewsQA	https://datasets.maluuba.com/NewsQA	News article-based comprehension QA
DeepMind	https://cs.nyu.edu/~kcho/DMQA/	News article-based QA
NLP QA	http://nlpprogress.com/	Many
TREC QA	https://trec.nist.gov/data/qa.html	Open-domain and closed-class questions
Yahoo! answers	https://webscope.sandbox.yahoo.com/	Many

capable of handling complex sentences. This will further give rise to research opportunities related in the realm of non-English CQAs.

In terms of approach, articles with traditional ML outnumbered those with DL. This is not surprising as DL is a more recently developed approach. In ML, researchers widely used the bag-of-word model and *tf-idf*, but probabilistic and statistical models were conspicuously rare. We found that very few studies have reported using the ensemble technique or using the multi-modal technique to resolve the issues of CQAs. Such techniques are encouraged going forward. Additionally, the use of DL in CQA research seems to be on an upward trajectory, and we welcome this trend. Also, people are nowadays not only limited to CQAs to get answers to their questions but also utilising microblogging sites such as Twitter and social networking sites such as Facebook. This sets the stage to investigate the extent to which ML and DL models that work on traditional CQAs apply to generic social media question answering.

7 | CONCLUSION, LIMITATIONS AND FUTURE SCOPE

This literature review paper has presented a summary of the studies conducted to address the issues of CQAs with computational methods including ML and DL. The studies were categorised into three modules: (i) the question module, (ii) the answer module, and (iii) the user module. Question unanswerability, duplication question detection, and closed question detection are the main issues in the question module. In the answer module, finding the best answer, answer quality prediction, and ranking of the answers are some of the pressing issues. The user module mainly includes two issues: finding expert users and reputation collectors. It was found that very few studies have used the ensemble technique or the multimodal technique.

In terms of dataset, some of the most widely studied platforms include YA, SO, and SE. However, we did not find any general model that was tested across different CQAs. The scope of most articles was confined to just one platform. Therefore, scholars interested in CQA research are encouraged to pursue cross-platform inquiry. CQAs that operate with different languages can also be compared.

This survey is, however, limited in the sense that it considered only English language-based research on CQA. Only studies with ML and DL were taken into consideration. Articles that did not have a clear focus on questions, answers and users were eliminated. The findings of the survey should be viewed in light of these limitations.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Adnan Gutub https://orcid.org/0000-0003-0923-202X

REFERENCES

1. Srba, I., et al.: Employing community question answering for online discussions in university courses: students' perspective. Comput. Educ. 135, 75–90 (2019)
2. Quan, X., Liu, W., Qiu, B.: Term weighting schemes for question categorization. IEEE Trans. Pattern Anal. Mach. Intell. 33(5), 1009–1021 (2010)
3. Dror, G., et al.: I want to answer; who has a question?: Yahoo! answers recommender system. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1109–1117. ACM (2011)
4. Roy, P.K., et al.: Identifying expert users on question answering sites. In: International Conference on Information Management & Machine Intelligence, pp. 285–291. Springer (2019)
5. Movshovitz-Attias, D., et al.: Analysis of the reputation system and user contributions on a question answering website: Stack overflow. In:

Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 886–893. ACM (2013)
6. Roy, P.K., et al.: Finding and ranking high-quality answers in community question answering sites. Global J. Flex. Syst. Manag. 19, 53–68 (2018)
7. Roy, P.K., Singh, J.P., Nag, A.: Finding active expert users for question routing in community question answering sites. In: International Conference on Machine Learning and Data Mining in Pattern Recognition, pp. 440–451. Springer (2018)
8. Kaysi, I., et al.: Users evaluation of rail systems in mass events: case study in mecca, Saudi Arabia, transportation research record. J. Transport. Res. Board. 2350(1), 111–118 (2013). http://doi.org/10.3141/2350-13
9. Al-Otaibi, N., Gutub, A.: Flexible stego-system for hiding text in images of personal computers based on user security priority. International Conference on Advanced Engineering Technologies (AET), pp. 250–256. Dubai UAE (2014)
10. Kolomiyets, O., Moens, M.-F.: A survey on question answering technology from an information retrieval perspective. Inf. Sci. 181(24), 5412–5434 (2011)
11. Allam, A.M.N., Haggag, M.H.: The question answering systems: a survey. Int. J. Res. Rev. Inf. Sci. 2(3), 1–12 (2012)
12. Gupta, P., Gupta, V.: A survey of text question answering techniques. Int. J. Comput. Appl. 53(4), 1–8 (2012)
13. Dwivedi, S.K., Singh, V.: Research and reviews in question answering system. Proc. Technol. 10, 417–424 (2013)
14. Bouziane, A., et al.: Question answering systems: survey and trends. Procedia Comput. Sci. 73, 366–375 (2015)
15. Kaysi, I., et al.: Users' Experience and Evaluation of New Southern Masha'er Metro Line in Makkah, Saudi Arabia. Transportation Research Board (TRB) Annual Meeting, Washington (2013)
16. Liu, Y., et al.: A survey on frameworks and methods of question answering. In: Information Science and Control Engineering (ICISCE), pp. 115–119. IEEE (2016)
17. Srba, I., Bielikova, M.: A comprehensive survey and classification of approaches for community question answering. ACM Trans. Web. 10(3), 181–186 (2016)
18. Wu, Q., et al.: Visual question answering: a survey of methods and datasets. Comput. Vis. Image Understand. 163, 21–40 (2017)
19. Diefenbach, D., et al.: Core techniques of question answering systems over knowledge bases: a survey. Knowl. Inf. Syst. 55(3), 529–569 (2018)
20. Soares, M.A.C., Parreiras, F.S.: A literature review on question answering techniques, paradigms and systems. J. King Saud Univ. Comput. Inf. Sci. 32(6), 635–646 (2020)
21. Mishra, A., Jain, S.K.: A survey on question answering systems with classification. J. King Saud Univ. Comput. Inf. Sci. 28(3), 345–361 (2016)
22. Ahmad, A., et al.: A survey on mining stack overflow: question and answering (q&a) community. Data Technologies and Applications. (2018)
23. Yang, Z., et al.: Expert recommendation in community question answering: a review and future direction. Int. J. Crowd Sci. 3(3), 348–372 (2019)
24. Khusro, S., Alam, A., Khalid, S.: Social question and answer sites: the story so far. program. 52(2), 170–192 (2017)
25. Yuan, S., et al.: Expert finding in community question answering: a review. Artif. Intell. Rev. 53(2), 843–874 (2020)
26. Altalhi, S., Gutub, A.: A survey on predictions of cyber-attacks utilizing real-time twitter tracing recognition. J. Ambient Intell. Hum. Comput. 12(11), 10209–10221 (2021)
27. Almutairi, S., Gutub, A., Al-Juaid, N.: Motivating teachers to use information technology in educational process within Saudi Arabia. Int. J. Technol. Enhanc. Learn. 12(2), 200–217 (2020)
28. Abu-Hashem, M., Gutub, A.: Efficient computation of Hash Hirschberg Protein Alignment Utilizing Hyper Threading Multi-Core Sharing Technology. CAAI Transactions on Intelligence Technology (2021). http://doi.org/10.1049/cit2.12070
29. Shambour, M.K., Gutub, A.: Personal privacy evaluation of smart devices applications serving hajj and umrah rituals. J. Eng. Res. (2021). http://doi.org/10.36909/jer.13199
30. Shambour, M.K., Gutub, A.: Progress of IoT research technologies and applications serving hajj and umrah. Arabian J. Sci. Eng. (2021). http://doi.org/10.1007/s13369-021-05838-7
31. Zhou, X., et al.: Recurrent convolutional neural network for answer selection in community question answering. Neurocomputing. 274, 8–18 (2018)
32. Li, M., et al.: Knowledge map construction for question and answer archives. Expert Syst. Appl. 141, 112923 (2020)
33. Roy, P.K., Singh, J.P.: Predicting closed questions on community question answering sites using convolutional neural network. Neural Comput. Appl. 32, 10555–10572, (2020)
34. Khan, A.U., et al.: Mmft-bert: Multimodal Fusion Transformer with Bert Encodings for Visual Question Answering (2020). arXiv preprint arXiv:2010.14095
35. Lukovnikov, D., Fischer, A., Lehmann, J.: Pretrained transformers for simple question answering over knowledge graphs. In: International Semantic Web Conference, pp. 470–486. Springer (2019)
36. McCarley, J., Chakravarti, R., Sil, A.: Structured Pruning of a Bert-Based Question Answering Model (2019). arXiv preprint arXiv:1910.06360
37. Yang, Z., et al.: Bert representations for video question answering. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 1556–1565 (2020)
38. Cao, Q., et al.: Deformer: Decomposing Pre-trained Transformers for Faster Question Answering (2020). arXiv preprint arXiv:2005.00697
39. Aly, S., et al.: Information gathering schemes for collaborative sensor devices. Procedia Comput. Sci. 32, 1141–1146 (2014)
40. Aly, S., et al.: Data dissemination and collection algorithms for collaborative sensor devices using dynamic cluster heads. Trends Appl. Sci. Res. 8(2), 55–72 (2013)
41. Kowsari, K., et al.: Text classification algorithms: a survey. Information. 10(4), 150 (2019)
42. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1097–1105 (2012)
43. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 9(8), 1735–1780 (1997)
44. Huang, Z., Xu, W., Yu, K.: Bidirectional LSTM-CRF models for sequence tagging. arXiv preprint arXiv:1508.01991 (2015)
45. Kim, Y.: Convolutional Neural Networks for Sentence classification (2014). arXiv preprint arXiv:1408.5882,
46. Kumar, A., et al.: A deep multi-modal neural network for informative twitter content classification during emergencies. Ann. Oper. Res. 1–32 (2020). http://doi.org/10.1007/s10479-020-03514-x
47. Mikolov, T., et al.: Efficient Estimation of Word Representations in Vector Space (2013). arXiv preprint arXiv:1301.3781
48. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: Empirical Methods in Natural Language Processing, pp. 1532–1543. EMNLP (2014)
49. Zhang, W.E., et al.: Detecting duplicate posts in programming QA communities via latent semantics and association rules. In: Proceedings of the 26th International Conference on World Wide Web, pp. 1221–1229. International World Wide Web Conferences Steering Committee (2017)
50. Wang, L., Zhang, L., Jiang, J.: Duplicate question detection with deep learning in stack overflow. IEEE Access. 8, 25964–25975 (2020)
51. Asaduzzaman, M., et al.: Answering questions about unanswered questions of stack overflow. In: Mining Software Repositories (MSR), 2013 10th IEEE Working Conference on, pp. 97–100. IEEE (2013)
52. Kim, S., et al.: Velocity-based modeling of physical interactions in dense crowds. Vis. Comput. 31, 541–555 (2015)
53. Curtis, S., et al.: Right of way: asymmetric agent interactions in crowds. Vis. Comput. 29(12), 1277–1292 (2013)
54. Abdelgawad, H., et al.: Microscopic modeling of large-scale pedestrian-vehicle conflicts in the city of Madinah, Saudi Arabia. J. Adv. Transport. 48(6), 507–525 (2014)

55. Hu, M., et al.: Read+ verify: machine reading comprehension with unanswerable questions. Proc. AAAI Conf. Artif. Intell. 33, 6529–6537 (2019)
56. Bhat, V., et al.: Effects of tag usage on question response time. Social Netw. Anal. Mining. 5(1), 24 (2015)
57. Mumtaz, S., Rodriguez, C., Benatallah, B.: Expert2vec: experts representation in community question answering for question routing. In: International Conference on Advanced Information Systems Engineering, pp. 213–229. Springer (2019)
58. Correa, D., Sureka, A.: Chaff from the wheat: characterization and modeling of deleted questions on stack overflow. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 631–642. ACM (2014)
59. Ponzanelli, L., et al.: Improving low quality stack overflow post detection. In: Software Maintenance and Evolution (ICSME), 2014 IEEE International Conference on, pp. 541–544. IEEE (2014)
60. Srba, I., Bielikova, M.: Why is stack overflow failing? preserving sustainability in community question answering. IEEE Softw. 33(4), 80–89 (2016)
61. Li, L., et al.: Answer quality characteristics and prediction on an academic q&a site: a case study on research gate. In: Proceedings of the 24th International Conference on World Wide Web, pp. 1453–1458 (2015)
62. Chua, A.Y., Banerjee, S.: So fast so good: an analysis of answer quality and answer speed in community question-answering sites. J. Am. Soc. Inf. Sci. Technol. 64(10), 2058–2068 (2013)
63. Gkotsis, G., et al.: It's all in the content: state of the art best answer prediction based on discretisation of shallow linguistic features. In: Proceedings of the 2014 ACM Conference on Web Science, pp. 202–210. ACM (2014)
64. Tondulkar, R., Dubey, M., Desarkar, M.S.: Get me the best: predicting best answerers in community question answering sites. In: Proceedings of the 12th ACM Conference on Recommender Systems, pp. 251–259. ACM (2018)
65. Zheng, W., Li, M.: The best answer prediction by exploiting heterogeneous data on software development q&a forum. Neurocomputing. 269, 212–219 (2017)
66. Hieber, F., Riezler, S.: Improved answer ranking in social question-answering portals. In: Proceedings of the 3rd International Workshop on Search and Mining Usergenerated Contents, pp. 19–26. ACM (2011)
67. Han, S., et al.: Answer ranking based on named entity types for question answering. In: Proceedings of the 11th International Conference on Ubiquitous Information Management and Communication, pp. 1–4 (2017)
68. Zhang, W., Deng, Y., Lam, W.: Answer ranking for product-related questions via multiple semantic relations modeling. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 569–578 (2020)
69. Kundu, D., Pal, R.K., Mandal, D.P.: Preference enhanced hybrid expertise retrieval system in community question answering services. Decis. Support Syst. 129, 113164 (2020)
70. Slag, R., de Waard, M., Bacchelli, A.: One-day flies on stackoverflow-why the vast majority of stackoverflow users only posts once. In: Mining Software Repositories (MSR), 2015 IEEE/ACM 12th Working Conference on, pp. 458–461. IEEE (2015)
71. Roy, P.K., et al.: Identifying reputation collectors in community question answering sites: exploring the dark side of social media. Int. J. Inf. Manag. 42, 25–35 (2018)
72. Li, B., et al.: Analyzing and predicting question quality in community question answering services. In: Proceedings of the 21st International Conference on World Wide Web, pp. 775–782. ACM (2012)
73. Zhang, Y., et al.: Multi-factor duplicate question detection in stack overflow. J. Comput. Sci. Technol. 30(5), 981–997 (2015)
74. Ahasanuzzaman, M., et al.: Mining duplicate questions of stack overflow. In: Mining Software Repositories (MSR), IEEE/ACM 13th Working Conference on, pp. 402–412. IEEE (2016)
75. Yang, L., et al.: Analyzing and predicting not-answered questions in community-based question answering services. AAAI. 11, 1273–1278 (2011)
76. Liang, D., et al.: Adaptive multi-attention network incorporating answer information for duplicate question detection. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 95–104 (2019)
77. Zhang, W.E., et al.: Feature analysis for duplicate detection in programming QA communities. In: International Conference on Advanced Data Mining and Applications, pp. 623–638. Springer (2017)
78. Chen, Z., et al.: Question retrieval for community based question answering via heterogeneous social influential network. Neurocomputing. 285, 117–124 (2018)
79. Silva, J., et al.: A 20% jump in duplicate question detection accuracy? replicating ibm teams experiment and finding problems in its data preparation. META. 20(4k), 1k (2018)
80. Godbole, A., Dalmia, A., Sahu, S.K.: Siamese Neural Networks with Random Forest for Detecting Duplicate Question Pairs (2018). arXiv preprint arXiv:1801.07288
81. Shah, D.J., et al.: Adversarial Domain Adaptation for Duplicate Question Detection (2018). arXiv preprint arXiv:1809.02255
82. Figueroa, A., G´omez-Pantoja, C., Neumann, G.: Integrating heterogeneous sources for predicting question temporal anchors across Yahoo! answers. Inf. Fusion. 50, 112–125 (2019)
83. Abric, D., et al.: Can duplicate questions on stack overflow benefit the software development community? In: Proceedings of the 16th International Conference on Mining Software Repositories, pp. 230–234. IEEE Press (2019)
84. Pal, A., Margatan, J., Konstan, J.: Question temporality: identification and uses. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, pp. 257–260 (2012)
85. Wu, F., et al.: Temporal interaction and causal influence in community-based question answering. IEEE Trans. Knowl. Data Eng. 29(10), 2304–2317 (2017)
86. Tao, K., et al.: Groundhog day: near duplicate detection on Twitter. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 1273–1284. ACM (2013)
87. Hoogeveen, D., et al.: Detecting misflagged duplicate questions in community question-answering archives. In: International Conference on Weblogs and Social Media, pp. 112–120. ICWSM (2018)
88. Muthmann, K., et al.: Near-duplicate detection for web-forums. In: Proceedings of the 2009 International Database Engineering & Applications Symposium, pp. 142–151. ACM (2009)
89. Zhang, W.E., et al.: Duplicate detection in programming question answering communities. ACM Trans. Internet Technol. 18(3), 37 (2018)
90. Xu, B., et al.: Predicting semantically linkable knowledge in developer online forums via convolutional neural network. In: Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering, pp. 51–62. ACM (2016)
91. Xu, B., et al.: Prediction of relatedness in stack overflow: deep learning vs. SVM: a reproducibility study. In: Proceedings of the 12th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ACM (2018)
92. Dror, G., Maarek, Y., Szpektor, I.: Will my question be answered? Predicting "question answerability" in community question-answering sites. ECML/PKDD 3, 499–514 (2013)
93. Correa, D., Sureka, A.: Fit or unfit: analysis and prediction of'closed questions' on stack overflow. In: Proceedings of the First ACM Conference on Online Social Networks, pp. 201–212. ACM (2013)
94. Xia, X., et al.: It takes two to tango: deleted stack overflow question prediction with text and meta features. In: Computer Software and Applications Conference (COMPSAC)IEEE 40th Annual. pp. 73–82. IEEE (2016)
95. Liu, J., Shen, H., Yu, L.: Question quality analysis and prediction in community question answering services with coupled mutual reinforcement. IEEE Trans. Services Comput. 10(2), 286–301 (2017)
96. Li, B., King, I.: Routing questions to appropriate answerers in community question answering services. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, pp. 1585–1588. ACM (2010)

97. Zhou, Y., et al.: Routing questions to the right users in online communities. In: IEEE 25th International Conference on Data Engineering, pp. 700–711. IEEE (2009)
98. Vakulenko, S., et al.: Question rewriting for conversational question answering. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining, pp. 355–363 (2021)
99. Zhou, T.C., Lyu, M.R., King, I.: A classification-based approach to question routing in community question answering. In: Proceedings of the 21st International Conference on World Wide Web, pp. 783–790. ACM (2012)
100. Chang, S., Pal, A.: Routing questions for collaborative answering in community question answering. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 13, pp. 494–501. Association for Computing Machinery, New York (2013)
101. Srba, I., Grznar, M., Bielikova, M.: Utilizing non-QA data to improve questions routing for users with low QA activity in CQA. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '15, pp. 129–136. Association for Computing Machinery, New York (2015)
102. Liang, Y., Introne, J.: Social roles, interactions and community sustainability in social q&a sites: a resource-based perspective. In: Proceedings of the 52nd Hawaii International Conference on System Sciences, pp. 2802–2811 (2019)
103. Li, Z., et al.: Personalized question routing via heterogeneous network embedding. AAAI. 33, 192–199 (2019)
104. Yue, A., et al.: Multi-feature based question–answerer model matching for predicting response time in CQA. Knowl. Base Syst. 182, 104794 (2019)
105. Wasim, M., et al.: Multi-label biomedical question classification for lexical answer type prediction. J. Biomed. Inf. 93, 1–13 (2019)
106. Melvin, Y.J., Gawade, S., Palivela, H.: Visual question answering using data mining techniques for skeletal scintigraphy in medical domain-VQADMSS. In: International Conference on Artificial Intelligence and Smart Systems (ICAIS), pp. 859–863. IEEE (2021)
107. Gupta, D., Suman, S., Ekbal, A.: Hierarchical deep multi-modal network for medical visual question answering. Expert Syst. Appl. 164, 113993 (2021)
108. Belhadi, A., et al.: Hybrid intelligent framework for automated medical learning. Expet Syst. 12, 1–25 (2021)
109. Bin-Hureib, E., Gutub, A.: Enhancing medical data security via combining elliptic curve cryptography and image steganography. Int. J. Comput. Sci. Netw. Security. 20(8), 1–8 (2020)
110. Samkari, H., Gutub, A.: Protecting medical records against cybercrimes within hajj period by 3-layer security. Recent Trends Inf. Technol. Appl. 2(3), 1–21 (2019)
111. Bin-Hureib, E., Gutub, A.: Enhancing medical data security via combining elliptic curve cryptography with 1-LSB and 2-LSB image steganography. Int. J. Comp. Sci. Netw. Security. 20(12), 232–241 (2020)
112. Espina, A., Figueroa, A.: Why was this asked? automatically recognizing multiple motivations behind community question-answering questions. Expert Syst. Appl. 80, 126–135 (2017)
113. Palomera, D., Figueroa, A.: Leveraging linguistic traits and semi-supervised learning to single out informational content across how-to community question-answering archives. Inf. Sci. 381, 20–32. (2017)
114. Harper, F.M., Moy, D., Konstan, J.A.: Facts or friends? Distinguishing informational and conversational questions in social q&a sites. In: Proceedings of the Sigchi Conference on Human Factors in Computing Systems, pp. 759–768 (2009)
115. Guy, I., et al.: Identifying informational vs. conversational questions on community question answering archives. In: Proceedings of the Eleventh Acm International Conference on Web Search and Data Mining, pp. 216–224 (2018)
116. Chua, A.Y., Banerjee, S.: Measuring the Effectiveness of Answers in Yahoo! Answers. Online Information Review (2015)
117. Sakai, T., et al.: Using graded relevance metrics for evaluating community QA answer selection. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, pp. 187–196. ACM (2011)
118. Zhu, Z., Bernhard, D., Gurevych, I.: A multi-dimensional model for assessing the quality of answers in social Q&A sites. Technische Universit¨ at Darmstadt (2009)
119. Rodrigo, A., Herrera, J., Penas, A.: The effect of answer validation on the performance of question-answering systems. Expert Syst. Appl. 116, 351–363 (2019)
120. Jeon, J., et al.: A framework to predict the quality of answers with non-textual features. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 228–235. ACM (2006)
121. Blooma, M.J., Chua, A.Y., Goh, D.H.-L.: A predictive framework for retrieving the best answer. In: Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 1107–1111. ACM (2008)
122. Agichtein, E., et al.: Finding high-quality content in social media. In: Proceedings of the 2008 International Conference on Web Search and Data Mining, pp. 183–194. ACM (2008)
123. Toba, H., et al.: Discovering high quality answers in community question answering archives using a hierarchy of classifiers. Inf. Sci. 261, 101–115. (2014)
124. Wu, H., et al.: An unsupervised approach for low-quality answer detection in community question-answering. In: International Conference on Database Systems for Advanced Applications, pp. 85–101. Springer (2017)
125. Zhang, Y., et al.: Understanding the formation mechanism of high-quality knowledge in social question and answer communities: a knowledge co-creation perspective. Int. J. Inf. Manag. 48, 72–84 (2019)
126. Hu, Z., et al.: A deep learning approach for predicting the quality of online health expert question-answering services. J. Biomed. Inf. 71, 241–253 (2017)
127. Suggu, S.P., et al.: Hand in glove: deep feature fusion network architectures for answer quality prediction in community question answering. In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp. 1429–1440 (2016)
128. Shah, C., Pomerantz, J.: Evaluating and Predicting Answer Quality in community QA. In: 33rd International ACM SIGIR conference on Research and Development in Information Retrieval, pp. 411–418. ACM (2010)
129. Lee, C.T., et al.: Model for voter scoring and best answer selection in community q&a servicesWI-IAT'09. In: IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technologies, pp. 116–123. IEEE (2009)
130. Blooma, M.J., Chua, A.Y.-K., Goh, D.H.-L.: Selection of the best answer in CQA services. In: Information Technology: New Generations (ITNG), International Conference on, pp. 534–539. IEEE (2010)
131. Burel, G., He, Y., Alani, H.: Automatic identification of best answers in online enquiry communities. In: Extended Semantic Web Conference, pp. 514–529. Springer (2012)
132. Tian, Q., Zhang, P., Li, B.: Towards predicting the best answers in community-based question-answering services. In: International Conference on Weblogs and Social Media, pp. 725–728. ICWSM (2013)
133. Chen, B.-C., et al.: Vote calibration in community question-answering systems. In: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 781–790. ACM (2012)
134. Elalfy, D., Gad, W., Ismail, R.: A hybrid model to predict best answers in question answering communities. Egypt. Inf. J. 19(1), 21–31 (2018)
135. Zhou, G., et al.: Exploring the existing category hierarchy to automatically label the newly-arising topics in CQA. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 1647–1651. ACM (2012)
136. Molino, P., Aiello, L.M., Lops, P.: Social question answering: textual, user, and network features for best answer prediction. ACM Trans. Inf. Syst. 35(1), 1–40 (2016)

137. Bhatt, G., Sharma, S., Raman, B.: Attentive Recurrent Tensor Model for community Question Answering (2018). arXiv preprint arXiv:1801.06792
138. Severyn, A., Moschitti, A.: Learning to rank short text pairs with convolutional deep neural networks. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 373–382. ACM (2015)
139. Dalip, D.H., et al.: Exploiting user feedback to learn to rank answers in q&a forums: a case study with stack overflow. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 543–552. ACM (2013)
140. Ginsca, A.L., Popescu, A.: User Profiling for Answer Quality Assessment in Q&A communities. Workshop on Data-Driven User Behavioral Modelling and Mining from Social Media, pp. 25–28. ACM (2013)
141. Stanovich, K.E.: Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. J. Educ. 189(1-2) 23–55 (2009)
142. Zhao, Z., et al.: Community-based question answering via asymmetric multi-faceted ranking network learning. In: AAAI, pp. 3532–3539 (2017)
143. Hu, B., et al.: Convolutional neural network architectures for matching natural language sentences. In: Advances in Neural Information Processing Systems, pp. 2042-2050. (2014)
144. Bouguessa, M., Dumoulin, B., Wang, S.: Identifying authoritative actors in question-answering forums: the case of Yahoo! answers. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 866–874. ACM (2008)
145. Chen, Q., et al.: Enhancing recurrent neural networks with positional attention for question answering. In: International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 993–996. ACM (2017)
146. Lukovnikov, D., et al.: Neural network-based question answering over knowledge graphs on word and character level. International Conference on World Wide Web, pp. 1211–1220. International World Wide Web Conferences Steering Committee (2017)
147. Nguyen, V.-T., Le, A.-C.: Deep neural network-based models for ranking question-answering pairs in community question answering systems. In: International Symposium on Integrated Uncertainty in Knowledge Modelling and Decision Making, pp. 179–190. Springer (2018)
148. Amancio, L., Dorneles, C.F., Dalip, D.H.: Recency and quality-based ranking question in CQAs: a stack overflow case study. Inf. Process. Manag. 58(4), 102552 (2021)
149. Xie, Y., et al.: Focusing attention network for answer ranking. In: The World Wide Web Conference, pp. 3384–3390 (2019)
150. Matsubara, Y., Vu, T., Moschitti, A.: Reranking for efficient transformer-based answer selection. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1577–1580 (2020)
151. Wu, W., et al.: Answer category-aware answer selection for question answering. IEEE Access. 9, 126357–126365 (2020)
152. Surdeanu, M., Ciaramita, M., Zaragoza, H.: Learning to rank answers on large online QA collections. In: Proceedings of ACL-08, pp. 719–727. HLT (2008)
153. Surdeanu, M., Ciaramita, M., Zaragoza, H.: Learning to rank answers to non-factoid questions from web collections. Comput. Ling. 37(2), 351–383 (2011)
154. Liu, M., Liu, Y., Yang, Q.: Predicting best answerers for new questions in community question answering. In: International Conference on Web-Age Information Management, pp. 127–138. Springer (2010)
155. Bordes, A., et al.: Large-Scale Simple Question Answering with Memory Networks. (2015) arXiv preprint arXiv:1506.02075
156. Zhang, J., Ackerman, M.S., Adamic, L.: Expertise networks in online communities: structure and algorithms. In: Proceedings of the 16th International Conference on World Wide Web, pp. 221–230. ACM (2007)
157. Aslay, C., et al.: Competition-based networks for expert finding. In: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1033–1036. ACM (2013)
158. Zhu, H., Chen, E., Cao, H.: Finding experts in tag based knowledge sharing communities. In: International Conference on Knowledge Science, Engineering and Management, pp. 183–195. Springer (2011)
159. Pal, A., et al.: Early detection of potential experts in question answering communities. In: International Conference on User Modeling, Adaptation, and Personalization, pp. 231–242. Springer (2011)
160. Pal, A., Harper, F.M., Konstan, J.A.: Exploring question selection bias to identify experts and potential experts in community question answering. ACM Trans. Inf. Syst. 30(2), 10–28 (2012)
161. Rafiei, M., Kardan, A.A.: A novel method for expert finding in online communities based on concept map and pagerank. Human Centric Comput. Inf. Sci. 5(1), 1–18 (2015)
162. Omidvar, A., Garakani, M., Safarpour, H.R.: Context based user ranking in forums for expert finding using wordnet dictionary and social network analysis. Inf. Technol. Manag. 15(1), 51–63 (2014)
163. Kundu, D., Mandal, D.P.: Formulation of a Hybrid Expertise Retrieval System in community Question Answering Services, pp. 1–15. Applied Intelligence (2018)
164. Wang, J., et al.: Convolutional neural networks for expert recommendation in community question answering. Sci. China Inf. Sci. 609(11), 110102 (2017)
165. Jurczyk, P., Agichtein, E.: Discovering authorities in question answer communities by using link analysis. In: Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, pp. 919–922. ACM (2007)
166. Page, L., et al.: The Pagerank citation Ranking: Bringing Order to the Web. Tech. Rep., Stanford (1999)
167. Liu, J., Song, Y.-I., Lin, C.-Y.: Competition-based user expertise score estimation. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 425–434. ACM (2011)
168. Rosen-Zvi, M., et al.: The author-topic model for authors and documents. In: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, pp. 487–494. AUAI Press (2004)
169. Cai, Y., Chakravarthy, S.: Expertise ranking of users in QA community. In: International Conference on Database Systems for Advanced Applications, pp. 25–40. Springer (2013)
170. Zhou, G., et al.: An empirical study of topic-sensitive probabilistic model for expert finding in question answer communities. Knowl. Base Syst. 66, 136–145 (2014)
171. Liu, Z., et al.: Toward personalized activity level prediction in community question answering websites. ACM Trans. Multimed Comput. Commun. Appl. 14, 1–15 (2018)
172. Ghasemi, N., Fatourechi, R., Momtazi, S.: User embedding for expert finding in community question answering. ACM Trans. Knowl. Discov. Data. 15(4), 1–16 (2021)
173. Liu, C., et al.: Identifying experts in community question answering website based on graph convolutional neural network. IEEE Access. 8, 137799–137811 (2020)
174. Fu, C.: Tracking user-role evolution via topic modeling in community question answering. Inf. Process. Manag. 56(6), 102075 (2019)
175. Saxena, A., Reddy, H.: Users roles identification on online crowdsourced q&a platforms and encyclopedias: a survey. J. Comput. Soc. Sci. 1–33 (2021). https://doi.org/10.1007/s42001-021-00125-9
176. Ford, D., Harkins, A., Parnin, C.: Someone like me: how does peer parity influence participation of women on stack overflow?. In: IEEE Symposium on Visual Languages and Human-Centric Computing, pp. 239–243. IEEE (2017)
177. Figueroa, A.: Male or female: what traits characterize questions prompted by each gender in community question answering? Expert Syst. Appl. 90, 405–413 (2017)
178. Ford, D.: Recognizing gender differences in stack overflow usage: Applying the Bechdel test. In: IEEE Symposium on Visual Languages and Human-Centric Computing, pp. 264–265. IEEE (2016)
179. Wang, Y.: Understanding the reputation differences between women and men on stack overflow. In: 25th Asia-Pacific Software Engineering Conference (APSEC), pp. 436–444. IEEE (2018)

180. Shambour, M.K., Gutub, A.: The Impact of Corona Disease on the Feelings of Twitter Users in the Two Holy Cities during the Hajj Season of 1442 AH. pp. 254–265. Umm Al-Qura University 21st Scientific Hajj Research Forum, Madinah, Saudi Arabia (2022). http://drive.uqu.edu.sa/up/profile/Books/4310001/book_path_FeMAOaNgOR.pdf
181. Figueroa, A., Timilsina, M.: What identifies different age cohorts in Yahoo! answers? Knowl. Base Syst. 228, 107278 (2021)
182. Sahu, A.K., Gutub, A.: Improving Grayscale Steganography to Protect Personal Information Disclosure within Hotel Services. Multimedia Tools and Applications (MTAP) (2022). http://doi.org/10.1007/s11042-022-13015-7
183. Gor, M., Webster, K., Boyd-Graber, J.: Toward deconfounding the influence of entity demographics for question answering accuracy. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 5457–5473 (2021)
184. Singh, A., et al.: AI based mobile edge computing for IoT: applications, challenges, and future scope. Arabian J. Sci. Eng. (2022). http://doi.org/10.1007/s13369-021-06348-2
185. Figueroa, A., Peralta, B., Nicolis, O.: Coming to grips with age prediction on imbalanced multimodal community question answering data. Information. 12(2), 48 (2021)
186. Gutub, A.: Adopting counting-based secret-sharing for e-Video Watermarking allowing Fractional Invalidation. Multimed. Tool. Appl. 81(7), 9527–9547 (2022)
187. Dong, L., et al.: Social influence on endorsement in social q&a community: moderating effects of temporal and spatial factors. Int. J. Inf. Manag. 61, 102396 (2021)
188. Shi, C., et al.: How learning effects influence knowledge contribution in online q&a community? a social cognitive perspective. Decis. Support Syst. 149, 113610 (2021)
189. Gutub, A.: Boosting Image Watermarking Authenticity Spreading Secrecy from counting-based Secret-sharing, CAAI Transactions on Intelligence Technology. IET (IEE) - Wiley (2022). http://doi.org/10.1049/cit2.12093
190. Li, K., Cheng, L., Teng, C.-I.: Voluntary sharing and mandatory provision: private information disclosure on social networking sites. Inf. Process. Manag. 57(1), 102128 (2020)
191. Al-Shaarani, F., Gutub, A.: Increasing participants using counting-based secret sharing via involving matrices and practical steganography. Arabian J. Sci. Eng. 47(2), 2455–2477 (2022)
192. Roy, P.K., Singh, J.P.: A tag2vec approach for questions tag suggestion on community question answering sites. In: International Conference on Machine Learning and Data Mining in Pattern Recognition, pp. 168–182. Springer (2018)

How to cite this article: Roy, P.K., et al.: Analysis of community question answering issues via machine-learning and deep learning: State-of-the-art review. CAAI Trans. Intell. Technol. 8(1), 95–117 (2023). https://doi.org/10.1049/cit2.12081