Journal of the American Medical Informatics Association, 28(10), 2021, 2287–2297
doi: 10.1093/jamia/ocab143
Advance Access Publication Date: 2 August 2021
Review

Review

A systematic review of automatic text summarization for biomedical literature and EHRs

Mengqian Wang,[1] **Manhua Wang**,[2] **Fei Yu**,[2,3] **Yue Yang**,[1] **Jennifer Walker**,[3] and **Javed Mostafa**[1,2,4]

[1]Carolina Health Informatics Program, University of North Carolina, Chapel Hill, North Carolina, USA, [2]iSchool, University of North Carolina, Chapel Hill, North Carolina, USA, [3]Health Sciences Library, University of North Carolina, Chapel Hill, North Carolina, USA, and [4]Biomedical Research Imaging Center, the School of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA

*Corresponding Author: Mengqian Wang, MA, Carolina Health Informatics Program, 335 S Columbia St., CB#7585, Chapel Hill, NC 27599-7585, USA (mengqian@email.unc.edu)

Received 6 February 2021; Revised 21 June 2021; Editorial Decision 22 June 2021; Accepted 24 June 2021

ABSTRACT

Objective: Biomedical text summarization helps biomedical information seekers avoid information overload by reducing the length of a document while preserving the contents' essence. Our systematic review investigates the most recent biomedical text summarization researches on biomedical literature and electronic health records by analyzing their techniques, areas of application, and evaluation methods. We identify gaps and propose potential directions for future research.
Materials and Methods: This review followed the PRISMA methodology and replicated the approaches adopted by the previous systematic review published on the same topic. We searched 4 databases (PubMed, ACM Digital Library, Scopus, and Web of Science) from January 1, 2013 to April 8, 2021. Two reviewers independently screened title, abstract, and full-text for all retrieved articles. The conflicts were resolved by the third reviewer. The data extraction of the included articles was in 5 dimensions: *input, purpose, output, method, and evaluation*.
Results: Fifty-eight out of 7235 retrieved articles met the inclusion criteria. Thirty-nine systems used single-document biomedical research literature as their input, 17 systems were explicitly designed for clinical support, 47 systems generated extractive summaries, and 53 systems adopted hybrid methods combining computational linguistics, machine learning, and statistical approaches. As for the assessment, 51 studies conducted an intrinsic evaluation using predefined metrics.
Discussion and Conclusion: This study found that current biomedical text summarization systems have achieved good performance using hybrid methods. Studies on electronic health records summarization have been increasing compared to a previous survey. However, the majority of the works still focus on summarizing literature.

Key words: automatic text summarization, machine learning, computational linguistics, biomedical and health sciences literature, electronic health records

INTRODUCTION

Information overload in the biomedical and health sciences

Biomedical and healthcare information has been increasing exponentially. The unprecedented amount of diverse textual data leads to information overload.[1] Information overload interferes with information seekers' information processing, diminishes their productivity, and prevents them from acquiring knowledge at the right time. For example, the records of chronically ill patients are difficult to present coherently.[2] A physician cannot read hundreds of clinical notes during a regular medical visit without any technological assistance. To achieve a comprehensive understanding of a patient's situation, a physician sometimes retrieves information from other sources, which worsens the scenario.

Studies have shown that the consequences of information overload can be fatal.[3–5] Information overload increases task demand and mental effort, which potentially impairs healthcare providers' understanding of patients' medical conditions and hinders the providers from making optimal medical decisions.

Reducing information overload with automatic text summarization

Interacting with computer-based systems and accessing textual data has become an integral part of healthcare providers' workflow.[6] As an application of Automatic Text Summarization (ATS) in the biomedical domain, biomedical text summarization (BTS) generates condensed and relevant representations of the source documents[7] computationally. It helps healthcare providers focus on the most valuable information, which benefits medical decision-making and enhances healthcare quality. Usability studies conducted with physicians for EHR summarization indicated the effectiveness of reading an automatically generated summary instead of the raw records.[8]

Background and history

New ideas and techniques are emerging in ATS.[9] ATS systems usually need modifications before becoming applicable in the biomedical domain. There has been a growing interest in BTS.[10] An informal literature survey conducted by Afantenos et al[10] identified 10 BTS studies published between 1999 and 2003. This survey focused on issues such as scaling to large collections of documents, personalization, portability to new subdomains, and the integration of summarization technology with practical applications. In 2014, Mishra et al published a systematic review on BTS, which covered articles collected from MEDLINE, IEEE Digital Library, and ACM Digital Library from 2000 to 2013. They identified the need to apply and evaluate text summarization in research and patient care settings.[11] Advances have been achieved since Mishra's review, and a more recent review is needed. Moradi & Ghadiri[8] provided a brief overview of the recent advances of BTS, but they did not take a systematic approach to examining the cited studies.

Study scope

This survey focused on EHR and biomedical literature summarizations because they are the major types of content that BTS research typically concentrates on. It should be noted, however, the major goals of summarizations differ significantly for the user groups that utilize EHRs and biomedical literature. The EHR summarization systems are designed to help healthcare providers make optimized clinical decisions by extracting and organizing patient-specific information and evidence. Whereas, the primary goal of the biomedical literature summarization is to provide researchers and other end users an effective and efficient way to read and interpret biomedical research evidence .

Therefore, this review aims at (1) systematically investigating the most recent research (ie, published between 2013 to 2021) of the BTS application on biomedical literature and EHR documents; (2) identifying techniques, areas of application, and evaluation methods in these studies; (3) identifying gaps and proposing potential directions for future research.

MATERIALS AND METHODS

This study followed Mishra et al's approach[11] based on IOM Standards for Systematic Reviews[12] and refined the review protocol for the research aims. We expanded the searching scope by adding 2 more databases (Web of Science and Scopus). The search formulation was modified with an expert review committee and a medical librarian. The search strategies and results can be found in Supplementary Appendix A. The following subsections describe the steps we followed to identify, screen, and extract relevant data from the included studies.

Data sources and searches

We searched PubMed, ACM Digital Library, Scopus, and Web of Science (containing IEEE). The literature search was from January 1st, 2013 to April 8th, 2021, aiming to update previous findings by Mishra et al[11] Important search terms included "biomedical text summarization," "medical summarization," "biomedical automatic summarization," etc. (see Supplementary Appendix A).

Article selection

We included an original research study if it developed and evaluated ATS methods in biomedical and health sciences and biomedical and health informatics. We excluded an article if (1) the summarized text was outside of the biomedical domain; (2) the article was not a research article, such as editorials and opinion papers; (3) the emphasis was on text summarization tools, but without an evaluation component; (4) the related techniques (eg, text mining) in the article were used to support text summarization but did not produce a summary at the end; (5) the article was not written in English; (6) the article was about image and multimedia summarization without a text summarization component; or (7) the study was included in Mishra et al.[11]

Title and abstract screening

Two reviewers (the first and second author) independently screened each article's title and abstract in Covidence.[13] Two reviewers had substantial agreement[14] on title-abstract screening (Cohen's kappa = 0.63). A third reviewer (the fourth author) helped resolve the conflicts. Also, both reviewers met to discuss any disagreement until they resolved discrepancies and achieved a consensus before moving to full-text screening.

Full-text screening

The same 2 reviewers independently screened the full text of 304 articles which passed the title-abstract screening. The agreement achieved in the full-text round of screening was much higher (Cohen's kappa = 0.86).[14] The third reviewer helped resolve the conflicts. The logic and the flow of the selection process are in Figure 1.

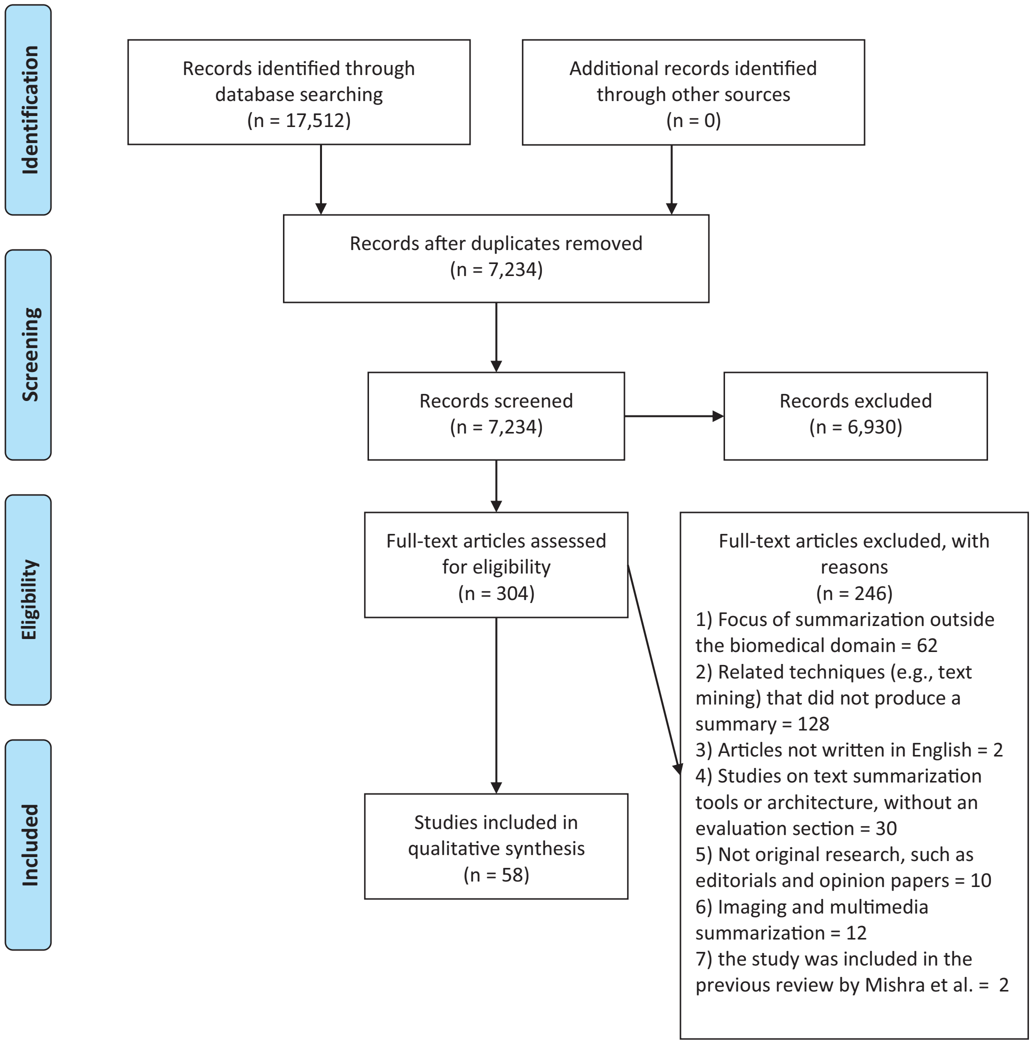

Figure 1. PRISMA diagram of the screening process.[15]

Data extraction and analysis

We developed a data extraction spreadsheet using Mani's standards.[16] Two reviewers (the first and second author) independently extracted the information from each included article according to the standards and resolved the conflicts with the third reviewer. Table 1 delineates the dimensions of our data extraction.

BTS input

Input characterizes the attributes of texts to be summarized, including (1) single vs multiple documents; (2) monolingual vs multilingual; (3) abstract vs full-text; (4) biomedical research literature vs EHR vs other document types (eg, medical news, clinical trial description, etc.).

BTS purpose

Purpose characterizes if the summarization system is (1) generic or user-oriented and if the system is (2) to facilitate clinical decision-making, to facilitate biomedical research, or to facilitate patient health information-seeking. A generic summarization system takes the predefined document(s) and generates a summary. In a user-oriented summarization system, the user provides a query or a list of parameters to customize the summaries.

"Healthcare providers" is the target user group for the clinical decision-making subcategory. The purpose of the summarization in the latter context was to digest clinical-related documents and deliver evidence-based knowledge. Some of the systems found were designed to aid patients in seeking health information. The subcategory of facilitating biomedical research is relatively broad. Studies summarizing biomedical literature without specifying a clinical purpose were counted as part of this group. It was found that some of the systems could belong to multiple subcategories. For example, the system discussed in Shree & Kiran[17] took EHRs as input and assisted with clinical decision support. At the same time, the latter system had features for protecting sensitive information, which contributes to biomedical research.[17]

BTS output

Output in this study focused on the following attributes of summarization.

(1) extractive vs abstractive

An extractive summarization system extracts sentences from the original text according to their importance. In contrast, an abstractive system extracts knowledge from the text and reconstructs them in a new piece of text.

Table 1. Dimensions of the data extraction

Classification scheme	Attributes
Input	1. Single vs Multiple Documents 2. Monolingual vs Multilingual 3. Abstract vs Full Text 4. Literature vs EHRs vs Others
Purpose	1. User-oriented vs Generic 2. To facilitate clinical decision-making vs To facilitate biomedical research vs To facilitate patient health information seeking
Output	1. Extractive vs Abstractive 2. Informative vs Indicative
Method	1. Statistics 2. Machine Learning (ML) 3. Computational Linguistics (CL) 4. Hybrid
Evaluation	1. Intrinsic vs Extrinsic 2. Quantitative vs Qualitative

(2) informative vs indicative

Indicative summaries provide users with a general idea of the input source content, but users need to refer back to the original text to understand the content. Informative summaries offer enough details so that users do not need to check the original content.

BTS method

The 3 categories of *method* are described below:

(1) Statistical

Using a rule-based statistical method, a researcher manually selects features and cues and makes calculations using a predefined formula. The features could be the position of the sentence, the important keywords that it contains, etc.

(2) Machine Learning (ML)

For a model defined up to some parameters, ML is the execution of a computer program to optimize the model's parameters using the training data or past experience.[18] ML uses statistics in building mathematical models because the core task is making inferences from a sample.[18] However, unlike statistical methods, the features of an ML method are selected automatically by an algorithm, and the parameters of a formula are not predefined. In our review, systems adopting ML usually combined other approaches. For example, Rouane et al[19] extracted key concepts from the sentences using MetaMap[20] and sorted concepts of each sentence into separate itemsets. K-means clustering, an unsupervised learning approach, was applied to group these itemsets automatically. Frequent itemsets were mined to weigh the sentences, and the higher-weighted sentences were extracted for the final summaries. In recent years, systems adapting pretrained word embeddings and trained using seq2seq models are thriving.[21–23] These systems are categorized as pure ML, as they do not need human-defined features or language analysis before training.

(3) Computational linguistics (CL)

CL investigates computational modeling of natural language. It includes simple applications, like word counting, and complicated ones, such as language generation. Our study categorizes a system as using CL techniques if it adopts text processing functions, such as extraction of lexical knowledge, lexical and structural disambiguation, grammatical inference, or robust parsing. For example, the system developed by Scott et al[24] took the Chronicle,[25] a knowledge-based semantic graph generated from raw clinical health records, as input. Chronicle retrieved subgraphs according to the type and extent of the summary requested by users. Their system determined the order of utterances using the knowledge retrieved[24] and generated sentences using a template-based grammar.[24]

(4) Hybrid

The hybrid method refers to using 2 or more methods from (1) to (3). For example, Gayathri et al[26] combined CL and statistical approaches. They extracted cue words using Medical Subject Headings (MeSH) and scored sentences based on the cue words' frequency as well as other features, such as sentence position and sentence length.

BTS evaluation

Evaluation could be Intrinsic or Extrinsic,[27] quantitative or qualitative.

(1) Intrinsic vs Extrinsic

An intrinsic evaluation method assesses the summaries internally according to specific criteria, such as scoring metrics (eg, ROUGE) or attributes, like readability, comprehensiveness, accuracy, and relevancy.

An extrinsic evaluation method assesses the summaries by applying them in a downstream task. It may measure users' efficiency, time, or accuracy when they complete a quiz using the automatically generated summaries.

(2) Quantitative vs Qualitative

In a quantitative evaluation, the metrics of measurement are clearly defined. The performance of the systems is evaluated based on their scores. On the contrary, qualitative evaluations do not have clear standards. The goal of a qualitative evaluation is usually exploration. In our review, most of the studies evaluated their systems quantitatively using predefined metrics, and some of the authors gave short qualitative analyses based on the quantitative results. We consider that a study adopts qualitative evaluation only if it has a separate section on qualitative analysis. Taking Moradi et al[28] as an example, besides the ROUGE metrics, it did a deep analysis on an example summary to bring out insights into their system.

RESULTS

A total of 58 publications were included. The characteristics and statistics of the included 58 studies (BTS systems) are summarized in

Tables 2 and 3 at the end of section 3 and Table B.1 in Supplementary Appendix B. The percentages in the following results are approximated.

BTS input

There were 39 (67%) studies/systems designed for single-document summarization, 18 (31%) studies designed for multiple-document summarization, and 1 (2%) study designed for both single- and multiple-document summarization. All included systems were designed for monolingual documents. Forty-seven (81%) studies took full text as input, 10 (17%) took abstract as input, and 1 (2%) took both. Forty (69%) systems summarized biomedical research literature, 11 (19%) systems summarized EHRs, and 7 (12%) systems summarized other biomedical documents. As indicated by the results above, it is popular to summarize biomedical scientific articles while using abstracts as reference summaries. The primary reason could be that scientific papers are easily accessible and have clear themes.

BTS purpose

Sixteen (28%) systems considered users' input (user-oriented) while 42 (72%) were generic. Seventeen (29%) systems were specifically designed to facilitate clinical decision-making, 5 (9%) systems were designed for helping with patient health information-seeking, and 34 (59%) focused on facilitating biomedical research. We classified 2 ([17,47]) as having multiple purposes.

BTS output

Forty-seven (81%) systems adopted extractive approaches, 10 (17%) used abstractive approaches, and 1 (2%) system applied both—then compared the results. Most included summarization systems output informative summaries, and 2 (3%) of the systems generated a summary containing informative and indicative content. Generating informative summaries by selecting salient sentences is 1 of the widely used strategies in BTS. As extractive summarization avoids the challenging language-generating task, it brings redundant information and incoherent sentences. We will further address this issue in section "Abstractive Approaches are under Development."

BTS method

Fifty-four (91%) systems applied hybrid methods, including the combination of Statistical and CL (22, 38%), the combination of ML and CL (4, 7%), the combination of Statistical and ML (1, 2%), and the combination of Statistical, ML, and CL (26, 45%). Four studies (7%) utilized ML only, and 1 (2%) applied CL only. The hybrid methods adopting CL were prevalent and efficient. We will further discuss these ideas in sections "Hybrid Methods," "CL Techniques," and "Syntactic Structures are Worth Further Investigation."

BTS evaluation

Fifty-one (88%) studies conducted an intrinsic evaluation, 3 (5%) studies conducted an extrinsic evaluation, and the remaining 4 (7%) studies conducted both types of evaluations. Thirty-eight (66%) studies conducted quantitative analyses only, and 2 (3%) studies used qualitative methods only. Eighteen (31%) studies included both metrics for quantitative evaluation and case analyses as qualitative evaluation. Some frequently used baselines include LexRank,[79] TextRank,[80] MEAD,[81] and the first/last several sentences.

Public availability

Fifty (86%) studies evaluated their systems using publicly available data, 2 (3%) of which included tests on private data as well. Eight (14%) studies linked to their source code or applications. More details can be found in Supplementary Table B.1 in Appendix B.

DISCUSSION

Our study is the first systematic review of BTS since 2014. Compared to the previous review,[11] we searched more literature databases and identified more relevant studies while keeping a similar scope for literature searching, study selection, and data extraction. Our results confirmed some previous findings[11] and showed unique BTS research trends in the most recent years.

State-of-the-art and improvement

Hybrid methods

The use of hybrid methods (91% in our review) has become the norm since 2013 (44% in Mishra et al[11]) confirming their observation that hybrid methods had great potential. The system proposed by Sarker et al[68] generated all possible 3-sentence combinations and then selected the combination having the highest ROUGE-L F-score as the ideal summary. They took advantage of supervised learning by using these ideal summaries to train their system and derived the final statistics of their summarization model. Besides the auto-generated statistics, they had relative sentence position and semantic types identified using UMLS. The system features also included manually composed formulas. Therefore, Sarker's system adopted a hybrid mode by combining CL, statistical, and ML. With the increasing availability of the pretrained word embeddings, the application of seq2seq ML models is growing. In reviewing the literature, it was found that, despite the increasing use of the latter ML models, the hybrid models remain in demand and they are frequently used in BTS.

CL techniques

Most hybrid methods (52 out of 53 articles) in this study included CL techniques. Knowledge-rich approaches that combine predefined domain knowledge are common in these methods. Mishra et al[11] believed that many publicly available knowledge resources, such as UMLS and MeSH and tools, such as MetaMap[20] and SemRep,[82] contributed to the high interest in CL. Including an expert-maintained domain knowledge database significantly improved the performance of various models. Gayathri et al[26] showed an example of a hybrid method combining CL and statistical approaches. They extracted cue words using MeSH terms and scored sentences based on the cue words' frequency and other features, such as sentence position. Gigioli et al[40] applied neural abstractive techniques[83] to the biomedical domain. Their abstractive summarization system was capable of generating novel summaries while considering domain knowledge. In addition, Gigioli et al explored maximum likelihood learning, reinforcement learning, and a mixed learning policy in their pointer–generator model. Both Gayathri's and Gigioli's studies compared their proposed methods with and without integrating biomedical knowledge. Their results indicated that integrating domain-related knowledge improved the performance of their models.

Public corpora

Increasingly, researchers have contributed to developing public corpora for BTS. For example, some studies (eg, Bavani et al[30]) used a

Table 2. Descriptive statistics of included studies based on study location and data extraction dimensions

Parameters	Category	Count	Percentage
Location	Multinational	10	17%
	Iran	7	12%
	USA	8	14%
	China	6	10%
	Israel	4	7%
	Australia	4	7%
	India	4	7%
	Algeria	3	5%
	Germany	3	5%
	South Korea	1	2%
	Thailand	1	2%
	Indonesia	1	2%
	Greece	1	2%
	Switzerland	1	2%
	Spain	1	2%
	Colombia	1	2%
	UK	1	2%
	Austria	1	2%
Input	Single document (SD)	39	67%
	Multiple documents (MD)	18	31%
	Single and Multiple documents (SD and MD)	1	2%
	Monolingual (Mono)	58	100%
	Multilingual (Multi)	0	0%
	Full Text (FT)	47	81%
	Abstract (Ab)	10	17%
	Full Text and Abstract (FT and Ab)	1	2%
	Biomedical Research Literature (Lit)	40	69%
	Electronic Health Record (EHR)	11	19%
	Other biomedical related documents	7	12%
Purpose	To facilitate clinical decision-making	17	29%
	To facilitate biomedical research	34	59%
	To facilitate patient health information seeking	5	9%
	To facilitate patient health information seeking/To facilitate clinical decision-making	1	2%
	To facilitate biomedical research/To support clinical decision-making	1	2%
	User-oriented (U)	16	28%
	Generic (G)	42	72%
Output	Informative	56	97%
	Informative and indicative	2	3%
	Extractive (Ext)	47	81%
	Abstractive (Abs)	10	17%
	Extractive and Abstractive (Ext and Abs)	1	2%
Method	Machine Learning (ML)	4	7%
	Statistical and Computational Linguistics (Stats and CL)	22	38%
	Computational Linguistics (CL)	1	2%
	Machine Learning and Statistical (ML and Stats)	1	2%
	Computational Linguistics and Machine learning (CL and ML)	4	7%
	Statistical, machine learning and Computational Linguistics (Stats, ML and CL)	26	45%
Evaluation	Intrinsic (I)	51	88%
	Extrinsic (E)	3	5%
	Intrinsic and Extrinsic (I and E)	4	7%
	Quantitative Evaluation (Quan)	38	66%
	Qualitative Evaluation (Qual)	2	3%
	Quantitative and Qualitative Evaluation (Quan and Qual)	18	31%
Data and Code	Data Publicly Available	48	83%
	Data Partially Available	2	3%
	Data not Publicly Available	6	10%
	Data Details not Mentioned	2	3%
	Source Code or Application Available	8	14%
	Source Code or Application not Available	51	88%

Table 3. Included studies by location and dimensions

Study	Location	Input	Purpose	Output	Evaluation	Public availability
Afzal, 2020[29]	Multinational	MD, Mono, Ab, Lit	CDM, U	Inf, Ext	I, Quan	DA
Bavani, 2016[30]	Australia	MD, Mono, Ab, Lit	CDM, U	Inf, Abs	I, Quan	DA
Bhaskoro, 2017[31]	Indonesia	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	
Bui, 2016[32]	USA	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Chiang, 2014[33]	Taiwan, China	MD, Mono, FT, Related	PHIS, G	Inf, Ext	I, Quan and Qual	DA
Cohan, 2018[34]	USA	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Conroy, 2018[35]	USA	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Davoodijam, 2021[36]	Iran	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Deng, 2020[22]	China	SD, Mono, FT, Related	BR, U	Inf, Ext	I, Quan and Qual	DA
Du, 2020[37]	China	SD, Mono, Ab, Lit	BR, G	Inf, Ext	I, Quan and Qual	DA
Dudko, 2017[38]	Multinational	MD, Mono, FT, EHR	CDM, G	Inf and Ind, Abs	I, Quan and Qual	DA
Gayathri, 2015[39]	India	SD, Mono, FT, Related	BR, G	Inf, Ext	I, Quan	DA
Gayathri, 2015[26]	India	SD, Mono, FT, Related	BR, G	Inf, Ext	I, Quan	DA
Gigioli, 2019[40]	USA	SD, Mono, Ab, Lit	BR, G	Inf, Abs	I, Quan and Qual	DA
Goldstein, 2013[41]	Israel	MD, Mono, FT, EHR	CDM, G	Inf, Abs	I, Qual	DA
Goldstein, 2015[42]	Israel	MD, Mono, FT, EHR	CDM, G	Inf, Abs	I, Qual	DA
Goldstein, 2016[43]	Israel	MD, Mono, FT, EHR	CDM, G	Inf, Abs	I and E, Quan	
Goldstein, 2017[44]	Israel	MD, Mono, FT, EHR	CDM, G	Inf, Abs	I and E, Quan and Qual	
Goodwin, 2020[21]	USA	SD, Mono, FT, Lit	PHIS, U	Inf, Abs	I, Quan	DA, CA
Gulden, 2019[45]	Germany	SD, Mono, FT, Related	BR, G	Inf, Ext	I, Quan, Qual	DA
Guo, 2013[46]	Multinational	SD, Mono, FT, Lit	BR, U	Inf, Ext	I and E, Quan and Qual	DA, CA
Kim, 2018[47]	South Korea	MD, Mono, FT, Lit	PHIS and CDM, U	Inf, Ext	I, Quan	DA
Lee, 2020[48]	USA	SD, Mono, FT, Lit	CDM, U	Inf and Ind, Ext	I, Quan and Qual	DA, CA
Liu, 2019[49]	China	MD, Mono, FT, Related	PHIS, G	Inf, Ext	I, Quan and Qual	DA
Lloret, 2013[50]	Spain	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan and Qual	DA
Malakasiotis, 2015[51]	Greece	MD, Mono, Ab, Lit	BR, U	Inf, Ext	I, Quan	DA
Mitrović, 2015[52]	Switzerland	SD, Mono, Ab and FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Moen, 2016[53]	Multinational	MD, Mono, FT, EHR	CDM, G	Inf, Ext	I, Quan	
Moradi, 2017[54]	Iran	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Moradi, 2018[55]	Iran	SD and MD, Mono, Ab and FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Moradi, 2018[56]	Iran	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Moradi, 2018[57]	Iran	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Moradi, 2019[58]	Austria	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA, CA
Moradi, 2020[28]	Multinational	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan and Qual	DA, CA
NasrAzadani, 2018[59]	Iran	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
NasrAzadani, 2018[60]	Iran	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Nguyen, 2013[61]	Multinational	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Parveen, 2015[62]	Germany	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Parveen, 2015[63]	Germany	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
PolepalliRamesh, 2015[64]	USA	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Puyana, 2013[65]	Colombia	MD, Mono, FT, EHR	CDM, U	Inf, Ext	I, Quan and Qual	
Rouane, 2019[66]	Algeria	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Rouane, 2019[19]	Algeria	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Rouane, 2020[67]	Algeria	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Sarker, 2013[68]	Australia	SD, Mono, Ab, Lit	CDM, U	Inf, Ext	I, Quan	DA
Sarker, 2016[69]	Australia	SD, Mono, Ab, Lit	CDM, U	Inf, Ext	I, Quan and Qual	DA
Scott, 2013[24]	UK	MD, Mono, FT, EHR	CDM, U	Inf, Ext	E, Quan and Qual	
Shree, 2020[17]	India	MD, Mono, FT, EHR	BR and CDM, G	Inf, Ext	I and E, Quan	
Sibunruang, 2018[70]	Thailand	SD, Mono, Ab, Lit	CDM, G	Inf, Ext	I, Quan	DA
Siranjeevi, 2020[71]	India	SD, Mono, FT, Lit	BR, G	Inf, Ext	I, Quan	DA
Song, 2020[23]	China	SD, Mono, Ab, Lit	BR, G	Inf, Abs	I, Quan	DA
Sotudeh, 2020[72]	USA	SD, Mono, FT, EHR	CDM, G	Inf, Abs	I, Quan	DA
Suominen, 2013[73]	Australia	MD, Mono, FT, EHR	CDM, G	Inf, Ext	E, Quan and Qual	
Ting, 2013[74]	Multinational	SD, Mono, FT, Lit	CDM, G	Inf, Ext	E, Quan and Qual	DA
Villa-Monte, 2019[75]	Multinational	SD, Mono, FT, Lit	BR, U	Inf, Ext	I, Quan	DA
Villa-Monte, 2020[76]	Multinational	SD, Mono, FT, Lit	BR, U	Inf, Ext	I, Quan	DA
Xu, 2016[77]	China	MD, Mono, Ab, Lit	PHIS, U	Inf, Ext	I, Quan	DA
Yin, 2014[78]	Multinational	MD, Mono, FT, Related	PHIS, U	Inf, Ext	I, Quan and Qual	DA

Abbreviations: Ab, abstract; Abs, abstractive; BR, to facilitate biomedical research; CA, code publicly available; CDM, to facilitate biomedical clinical decision-making; DA, data publicly available (including partial); E, extrinsic; EHR, electronic health record; Ext, extractive; FT, full text; G, generic; I, intrinsic; Ind, indicative; Inf, informative; Lit, Literature; MD, multiple documents; ML, multilingual; Mono, monolingual; PHIS, to facilitate patient health information seeking; Qual, qualitative evaluation; Quan, quantitative evaluation; SD, single document; U, user-oriented.

specialized evidence-based medicine corpus—which was gathered and annotated by Mollá et al[84] for the sole purpose of BTS. This corpus was sourced from the Clinical Inquiries section in the *Journal of Family Practice*, consisting of 456 clinical queries with 1396 bottom-line evidence-based answers. For each bottom-line answer, there existed a set of detailed justifications. Each detailed justification was in turn associated with at least 1 source document. Thus, this dataset can be reused for either multi-document or single-document summarization tasks.

Experiments in real-world settings and usability tests with physicians and patients

Among the studies designed for facilitating clinical decision-making by summarizing EHRs, 7 of them assessed their systems in real-world settings by conducting usability tests. Goldstein et al[43] evaluated their system through intrinsic and extrinsic usability tests. There were 3 components in their evaluation session: (1) relative completeness by requesting physicians to tag if the missing items in a generated summary were essential; (2) quality analysis by checking readability, comprehensiveness, clinical course, and continuity of care; (3) functional analysis based on the correctness and time of physicians answering 5 basic clinical decision questions. On average, physicians answered the questions 40% faster ($P < .001$) when using a system-generated letter than when using a physician-composed letter. Considering the correctness of the answers, they found that for 4 out of 5 questions, physicians did equally well or significantly better ($P < .005$) when using the system-generated letter. Moen et al[53] tried to validate the reliability of automatic evaluations on EHR summarization. They tested Spearman's rank correlation coefficient between the scores assessed by domain experts and by ROUGE metrics.

For those studies designed for facilitating patients' information-seeking, 2 of them conducted usability tests. In Liu et al,[49] the participants rated their satisfaction regarding the usefulness and representativeness of the summaries generated by the proposed and the baseline system. And in Yin et al,[78] 3 users compared the generated summaries and the results by a search engine. Both of them found the proposed systems satisfied users.

All usability tests with users (either physicians or patients) were conducted using surveys or interviews with predefined metrics. It needs to be noted that, in some of the studies, the authors manually analyzed examples. These were referred to as "Additional qualitative evaluation" and "Preliminary evaluation." More details can be found in Supplementary Appendix B Table B.1.

Automatic text summarization of EHRs

As shown in the studies by Goldstein et al[43] and Scott et al,[24] a summary of health records helped physicians save a significant amount of time processing patients' information and improved clinical decision accuracy. Nineteen percent of the included studies provided solutions to summarizing EHRs. This is a significant increase from the 9% identified by Mishra et al's review.[11]

Worldwide development

Text summarization in the biomedical domain has become a research focus worldwide. The included studies were conducted by researchers from 17 different countries (Table 3), 6 more countries than identified previously.[11] The USA and Australia remain among the top producers of BTS research, but researchers from several other countries (Iran, China, Israel, and India) produced 4 or more studies included in this review. The trend of multinational collaboration in BTS research has been persistent. Besides, systems have been developed for text summarization in different languages other than English.[33]

Gaps and challenges

Despite research progress made in recent years, this study identified a few gaps and challenges.

Syntactic structures are worth further investigation

When using CL techniques, most approaches emphasized semantic knowledge. Syntactic features that compose an essential part of CL are frequently ignored. However, combining syntactic features can increase concept and relation extraction accuracy.[85] As for information extraction, understanding the role of the extracted piece in the sentence is crucial. Integrating syntactic features in BTS systems is worth further investigation. In those seq2seq models where no features need to be specified, we believe that the order information embedded in the hidden layers is potentially beneficial.

Abstractive approaches are under development

The majority of the systems are extractive. Abstractive summarization has extra challenges of natural language generation. Parveen et al[62] found that the machine-generated abstractive summaries might have readability issues even if they cover all essential information. However, extractive summaries might contain redundant information that impacts the summary quality. As the space is limited, redundancy may result in a core information deficiency. Therefore, investigations are needed for developing intuitive, efficient, and context-sensitive abstractive summarization systems.

Challenges of EHR summarization

BTS for EHRs in clinical settings is still under development. This study observed several challenges in the EHR summarizations. First, free notes containing inconsistent abbreviations, incomplete sentences, and unclear implications increased the difficulty of text summarization. CL tools built on semantic knowledge databases are widely used to deal with these problems. However, the databases need intensive maintenance and are not always up-to-date. Second, we still lack universal datasets for EHR summarization. Concerning patients' privacy and security, researchers have difficulty accessing real patient records or acquiring the corresponding gold standard summaries generated by human professionals. Goldstein et al[43] conducted their experiment on MIMIC, an openly available dataset comprising deidentified health data.[86] Due to the lack of gold standard summaries, they used the discharge summary as an alternative, potentially impacting their system's overall performance and the user experience. Third, there has been no widespread adoption of BTS in clinical settings. Deployment is often hindered by a wide variety of established commercial EHR systems. The lack of rigorous evaluation is another large barrier for translating research into clinical practices.[87] Therefore, developing universal and high-quality EHR summarization datasets is vital for research and actual deployment.

LIMITATIONS

First of all, due to the scope of this study, some commercial BTS systems may have been inadvertently overlooked. Second, a meta-analysis comparing the performance of different approaches was not possible due to the heterogeneity of the evaluation methods in the studies reviewed. ROUGE metrics, although prevalent, are not the only option. The lack of a widely used and standardized dataset brought challenges for comparing different systems. Third, as our data screening, extraction, and analysis were guided by Mani's framework,[16] there might be additional dimensions and trends outside of Mani's framework not included in this study. For example, since most of the systems utilized CL techniques and extractive approaches, these 2 dimensions could be further divided into more granular classes in future reviews. Finally, relevant studies published in languages other than English were excluded.

CONCLUSION

This study systematically reviewed the latest research publications of text summarizations of biomedical literature and EHRs. The review covered articles published from 2013 to April 8, 2021, immediately following the last published systematic review on the same topic. Our findings demonstrate that the current BTS systems had achieved good performance using hybrid methods. It was found that CL techniques, especially knowledge-rich approaches, deliver positive outcomes. However, as essential components of CL techniques, the power of syntactic parsing and features have not been fully leveraged in BTS systems. Last but not least, most BTS systems were still designed for summarizing biomedical research literature rather than EHRs.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

AUTHOR CONTRIBUTIONS

MeW, MaW, and YY are the reviewers of the searched studies. MeW, FY, and JW contributed to the searching and deduplicating of the studies. MeW, FY, and JM contributed to the writing and editing of the paper.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

DATA AVAILABILITY STATEMENT

The data underlying this article are available in the article and in its online supplementary material .

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Stead WW, Lin HS, eds. *Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions*. Washington, DC: National Academies Press; 2009.
2. Christensen T, Grimsmo A. Instant availability of patient records, but diminished availability of patient information: a multi-method study of GP's use of electronic patient records. *BMC Med Inform Decis Mak* 2008; 8 (1): 12–8.
3. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectibility of man. *N Engl J Med* 1976; 295 (24): 1351–5.
4. McDonald CJ, Callaghan FM, Weissman A, *et al*. Use of internist's free time by ambulatory care electronic medical record systems. *JAMA Intern Med* 2014; 174 (11): 1860–3.
5. Karsh BT, Holden RJ, Alper SJ, *et al.* A human factors engineering paradigm for patient safety: designing to support the performance of the healthcare professional. *Qual Saf Health Care* 2006; 15 (SUPPL 1): i59–65.
6. Mazur LM, Mosaly PR, Moore C, *et al*. Toward a better understanding of task demands, workload, and performance during physician-computer interactions. *J Am Med Informatics Assoc* 2016; 23 (6): 1113–20.
7. Torres-MorenoJ-M. *Automatic Text Summarization*. Hoboken, NJ: John Wiley & Sons; 2014. doi:10.1002/9781119004752
8. Moradi M, Ghadiri N. Text summarization in the biomedical domain. *arXiv Prepr. arXiv1908.02285* 2019.
9. Allahyari M, Pouriyeh S, Assefi M, *et al*. Text summarization techniques: a brief survey. *arXiv Prepr. arXiv1707.02268* 2017; doi:10.14569/ijacsa.2017.081052
10. Afantenos S, Karkaletsis V, Stamatopoulos P. Summarization from medical documents: a survey. *Artif Intell Med* 2005; 33 (2): 157–77.
11. Mishra R, Bian J, Fiszman M, *et al.* Text summarization in the biomedical domain: a systematic review of recent research. *J Biomed Inform* 2014; 52: 457–67.
12. Eden J , Levit L, Berg A, *et al*., eds. *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: National Academies Press; 2011.
13. Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. www.covidence.org Accessed April 8, 2021.
14. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012; 22 (3): 276–82.
15. Moher D, Liberati A, Tetzlaff J, *et al*.; The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med* 2009; 6 (7): e1000097.
16. Mani I. *Automatic Summarization*. Amsterdam: John Benjamins Publishing; 2001.
17. Shree ANR, Kiran P. Sensitivity Context Aware Privacy Preserving Text Document Summarization. In: proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology, ICECA*; 5–7 Nov 2020; Tamil Nadu, India*. doi:10.1109/ICECA49313.2020.9297415
18. Alpaydin E. *Introduction to Machine Learning*. Cambridge, MA: MIT Press; 2020.
19. Rouane O, Belhadef H, Bouakkaz M. Combine clustering and frequent itemsets mining to enhance biomedical text summarization. *Expert Syst Appl* 2019; 135: 362–73.
20. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. *J Am Med Informatics Assoc* 2010; 17 (3): 229–36. doi:10.1136/jamia.2009.002733
21. Goodwin T, Savery M, Demner-Fushman D. Towards zero-shot conditional summarization with adaptive multi-task fine-tuning. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*; 16–20 Nov 2020; Virtual. doi:10.18653/v1/2020.findings-emnlp.289
22. Deng Y, Zhang W, Li Y, *et al.* Bridging hierarchical and sequential context modeling for question-driven extractive answer summarization. In: *SIGIR 2020—proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*; 23–30 Jul 2020; Xi'an, China. doi:10.1145/3397271.3401208.
23. Song G, Wang Y. A hybrid model for medical paper summarization based on COVID-19 open research dataset. In: *4th International Conference on Computer Science and Artificial Intelligence*; 23–25 Jul 2020; Stockholm, Sweden. doi:10.1145/3445815.3445824
24. Scott D, Hallett C, Fettiplace R. Data-to-text summarisation of patient records: using computer-generated summaries to access patient histories. *Patient Educ Couns* 2013; 92 (2): 153–9. doi:10.1016/j.pec.2013.04.019

25. Harkema H, Roberts I, Gaizauskas R, *et al.* Information extraction from clinical records. In: *Proceedings of the 4th UK e-Science All Hands Meeting*; 2005: 19–22.
26. Gayathri P, Jaisankar N. An efficient medical document summarization using sentence feature extraction and ranking. *Indian J Sci Technol* 2015; 8 (33): 1–8. doi:10.17485/ijst/2015/v8i33/71257.
27. Jones KS, Galliers JR, eds. *Evaluating Natural Language Processing Systems*. Berlin: Springer; 1995. doi:10.1007/BFb0027470
28. Moradi M, Dashti M, Samwald M. Summarization of biomedical articles using domain-specific word embeddings and graph ranking. *J Biomed Inform* 2020; 107: 103452.
29. Afzal M, Alam F, Malik KM, *et al.* Clinical context–aware biomedical text summarization using deep neural network: model development and validation. *J Med Internet Res* 2020; 22 (10): e19810.
30. Bavani ES, Ebrahimi M, Wong R, *et al.* Appraising UMLS coverage for summarizing medical evidence. In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers;* 11–16 Dec 2016; Osaka, Japan.
31. Bhaskoro SB, Akbar S, Supangkat SH, *et al.* Extracting important sentences for public health surveillance information from Indonesian medical articles. In: *2017 International Conference on Ict for Smart Society*; 18–19 Sep 2017: 1–7; Tangerang, Indonesia.
32. Bui DDA, Del Fiol G, Hurdle JF, *et al.* Extractive text summarization system to aid data extraction from full text in systematic review development. *J Biomed Inf* 2016; 64: 265–72.
33. Chiang CL, Chen SY, Cheng PJ, *et al.* Summarizing search results with community-based question answering. In: *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*; 11–14 Aug 2014; Warsaw, Poland. doi:10.1109/WI-IAT.2014.41
34. Cohan A, Goharian N. Scientific document summarization via citation contextualization and scientific discourse. *Int J Digit Libr* 2018; 19 (2–3): 287–303.
35. Conroy JM, Davis ST. Section mixture models for scientific document summarization. *Int J Digit Libr* 2018; 19 (2–3): 305–22.
36. Davoodijam E, Ghadiri N, Lotfi Shahreza M, *et al.* MultiGBS: a multi-layer graph approach to biomedical summarization. *J Biomed Inform* 2021; 116: 103706.
37. Du Y, Li Q, Wang L, *et al.* Biomedical-domain pre-trained language model for extractive summarization. *Knowledge-Based Syst* 2020; 199: 105964.
38. Dudko A, Endrjukaite T, Kiyoki Y. Medical documents processing for summary generation and keywords highlighting based on natural language processing and ontology graph descriptor approach. In: *Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services*; 4–6 Dec 2017; Salzburg, Austria. doi:10.1145/3151759.3151784
39. Gayathri P, Jaisankar N. Towards an efficient approach for automatic medical document summarization. *Cybern Inf Technol* 2015; 15 (4): 78–91.
40. Gigioli P, Sagar N, Rao A, *et al.* Domain-aware abstractive text summarization for medical documents. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*; 3–6 Dec 2018: 2338–43; Madrid, Spain. doi:10.1109/BIBM.2018.8621539
41. Goldstein A, Shahar Y. Implementation of a system for intelligent summarization of longitudinal clinical records. *Process Support Knowl Represent Heal Care*; 2013: 68–82. doi:10.1007/978-3-319-03916-9_6
42. Goldstein A, Shahar Y. Generation of natural-language textual summaries from longitudinal clinical records. *Stud Heal Technol Inf* 2015; 216: 594–8.
43. Goldstein A, Shahar Y. An automated knowledge-based textual summarization system for longitudinal, multivariate clinical data. *J Biomed Inform* 2016; 61: 159–75.
44. Goldstein A, Shahar Y, Orenbuch E, *et al.* Evaluation of an automated knowledge-based textual summarization system for longitudinal clinical data, in the intensive care domain. *Artif Intell Med* 2017; 82: 20–33.
45. Gulden C, Kirchner M, Schuttler C, *et al.* Extractive summarization of clinical trial descriptions. *Int J Med Inform* 2019; 129: 114–21.
46. Guo Y, Silins I, Stenius U, *et al.* Active learning-based information structure analysis of full scientific articles and two applications for biomedical literature review. *Bioinformatics* 2013; 29 (11): 1440–7.
47. Kim GW, Lee DH. Personalised health document summarisation exploiting Unified Medical Language System and topic-based clustering for mobile healthcare. *J Inf Sci* 2018; 44 (5): 619–43.
48. Lee EK, Uppal K. CERC: an interactive content extraction, recognition, and construction tool for clinical and biomedical text. *BMC Med Inform Decis Mak* 2020; 20 (S14): 1–14. doi:10.1186/s12911-020-01330-8
49. Liu YH, Song X, Chen SF. Long story short: finding health advice with informative summaries on health social media. *Aslib J Inf Manag* 2019; 71 (6): 821–40. doi:10.1108/AJIM-02-2019-0048
50. Lloret E, Romá-Ferri MT, Palomar M. COMPENDIUM: A text summarization system for generating abstracts of research papers. *Data Knowl Eng* 2013; 88: 164–75.
51. Malakasiotis P, Archontakis E, Androutsopoulos I, *et al.* Biomedical Question-Focused Multi-Document Summarization: ILSP and AUEB at BioASQ3. *CLEF (Working Notes)* 2015.
52. Mitrović S, Müller H. Summarizing Citation Contexts of Scientific Publications. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 2015; 9283: 154–65. doi: 10.1007/978-3-319-24027-5_13
53. Moen H, Peltonen LM, Heimonen J, *et al.* Comparison of automatic summarisation methods for clinical free text notes. *Artif Intell Med* 2016; 67: 25–37.
54. Moradi M, Ghadiri N. Quantifying the informativeness for biomedical literature summarization: an itemset mining method. *Comput Methods Programs Biomed* 2017; 146: 77–89.
55. Moradi M. CIBS: a biomedical text summarizer using topic-based sentence clustering. J Biomed Inform2018; 88: 53–61.
56. Moradi M. Frequent itemsets as meaningful events in graphs for summarizing biomedical texts. In: *8th International Conference on Computer and Knowledge Engineering (ICCKE)*; 25–26 Oct 2018: 135–40; Mashhad, Iran. doi:10.1109/ICCKE.2018.8566651
57. Moradi M, Ghadiri N. Different approaches for identifying important concepts in probabilistic biomedical text summarization. *Artif Intell Med* 2018; 84: 101–16.
58. Moradi M, Dorffner G, Samwald M. Deep contextualized embeddings for quantifying the informative content in biomedical text summarization. *Comput Methods Programs Biomed* 2020; 184: 105117.
59. Nasr Azadani M, Ghadiri N. Evaluating different similarity measures for automatic biomedical text summarization. In: *International Conference on Intelligent Systems Design and Applications*; 14–16 Dec 2017: 305–14; Delhi, India. doi:10.1007/978-3-319-76348-4_30.
60. Nasr Azadani M, Ghadiri N, Davoodijam E. Graph-based biomedical text summarization: an itemset mining and sentence clustering approach. *J Biomed Inform* 2018; 84: 42–58.
61. Nguyen DT, Leveling J. Exploring domain-sensitive features for extractive summarization in the medical domain. In: *International Conference on Application of Natural Language to Information Systems*; 19–21 Jun 2013: 90–101; Manchester, UK. doi:10.1007/978-3-642-38824-8_8
62. Parveen D, Strube M. Integrating importance, non-redundancy and coherence in graph-based extractive summarization. In: *Twenty-Fourth International Joint Conference on Artificial Intelligence*; 25–31 Jul 2015: 1298–304; Buenos Aires, Argentina. doi: 10.5555/2832415.2832430
63. Parveen D, Ramsl HM, Strube M. Topical coherence for graph-based extractive summarization. In: *Proceedings of the 2015 Conference on*

Empirical Methods in Natural Language Processing; 17–21 Sep 2015: 1949–54; Lisbon, Portugal. doi: 10.18653/v1/D15-1226

64. Polepalli Ramesh B, Sethi RJ, Yu H. Figure-associated text summarization and evaluation. *PLoS One* 2015; 10 (2): e0115671.
65. Puyana CG, Quimbaya AP. *GReAT: A Model for the Automatic Generation of Text Summaries*. In: *International Conference on Enterprise Information Systems*; 4–7 Jul 2013: 280–8; Angers Loire Valley, France.
66. Rouane O, Belhadef H, Bouakkaz M, *et al.* A New Biomedical Text Summarization Method Based on Sentence Clustering and Frequent Itemsets Mining. In: *International conference on the Sciences of Electronics, Technologies of Information and Telecommunications*; 27–29 Dec 2020: 144–52; Sanya, China. doi:10.1007/978-3-030-21005-2_14
67. Rouane O, Belhadef H, Bouakkaz M. *Word embedding-based biomedical text summarization. In:* Advances in Intelligent Systems and Computing*;* 2020. doi:10.1007/978-3-030-33582-3_28.
68. Sarker A, Mollá D, Paris C, *et al.* An approach for query-focused text summarisation for evidence-based medicine. In: *Conference on Artificial Intelligence in Medicine in Europe;* 29 May–1 Jun 2013: 295–304; Murcia, Spain. doi:10.1007/978-3-642-38326-7_41.
69. Sarker A, Mollá D, Paris C. Query-oriented evidence extraction to support evidence-based medicine practice. *J Biomed Inform* 2016; 59: 169–84.
70. Sibunruang C, Polpinij J. Finding clinical knowledge from MEDLINE abstracts by text summarization technique. In: *2018 International Conference on Information Technology (InCIT)*; 24–25 Oct 2018; Khon Kaen, Thailand. doi:10.23919/INCIT.2018.8584867
71. Siranjeevi H, Venkatraman S, Krithivasan K. Text summarization by hybridization of hypergraphs and hill climbing technique. In: *Advances in Intelligent Systems and Computing*. 2020. doi:10.1007/978-981-15-1286-5_28.
72. Sotudeh S, Goharian N, Filice RW. Attend to medical ontologies: content selection for clinical abstractive summarization. *In:* arXiv. 2020: 1899–905. doi:10.18653/v1/2020.acl-main.172.
73. Suominen H, Hanlen L. 10.1145/2537734.2537739. Visual summarisation of text for surveillance and situational awareness in hospitals. In: *Proceedings of 18th Australas Document Computing Symposium; 5*–6 Dec 2013: 89–96; New York, United States. doi:10.1145/2537734.2537739
74. Ting SL, See-To EWK, Tse YK. Web information retrieval for health professionals. *J Med Syst* 2013; 37 (3): 9946.
75. Villa-Monte A, Lanzarini L, Bariviera AF, *et al.* User-oriented summaries using a PSO based scoring optimization method. *Entropy* 2019; 21 (6): 617.
76. Villa-Monte A, Lanzarini L, Corvi J, *et al.* Document summarization using a structural metrics based representation. *J Intell Fuzzy Syst* 2020; 38 (5): 5579–88.
77. Xu B, Lin H, Hao H, *et al.* Generating User-Oriented Text Summarization Based on Social Networks Using Topic Models. In: *Chinese National Conference on Social Media Processing*; 29–30 Oct 2016: 186–93; Nanjing, China. 10.1007/978-981-10-2993-6_16
78. Yin Y, Zhang Y, Liu X, *et al.* HealthQA: A Chinese QA summary system for smart health. *LNCS* 2014; 8549: 51–62.
79. Erkan G, Radev DR. LexRank: Graph-based lexical centrality as salience in text summarization. *J Artif Intell Res* 2004; 22: 457–79. doi:10.1613/jair.1523
80. Mihalcea R, Tarau P. TextRank: Bringing order into texts. In: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing; 25–26 Jul 2004: 404–11; Barcelona, Spain.
81. Radev DR, Jing H, Styś M, *et al.* Centroid-based summarization of multiple documents. *Inf Process Manag* 2004; 40 (6): 919–38. doi:10.1016/j.ipm.2003.10.006
82. Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *J Biomed Inform* 2003; 36 (6): 462–77. doi:10.1016/j.jbi.2003.11.003.
83. Rush AM, Chopra S, Weston JA. Neural attention model for abstractive sentence summarization. *arXiv Prepr arXiv150900685* 2015.
84. Mollá D, Santiago-Martínez ME, Sarker A, *et al.* A corpus for research in text processing for evidence based medicine. *Lang Resources & Evaluation* 2016; 50 (4): 705–27.
85. Geng ZQ, Chen GF, Han YM, *et al.* Semantic relation extraction using sequential and tree-structured LSTM with attention. *Inf Sci (Ny)* 2020; 509: 183–92.
86. Johnson AEW, Pollard TJ, Shen L, *et al.* MIMIC-III, a freely accessible critical care database. *Sci Data* 2016; 3 (1): 1–9.
87. Pivovarov R. *Electronic Health Record Summarization over Heterogeneous and Irregularly Sampled Clinical Data*. ProQuest Diss. Theses; 2016. doi:10.7916/D89W0F6V.