An Intelligent Literature Review: an Inductive Approach to define Machine Learning Applications in the clinical domain

Renu Sabharwal (Renu.Sabharwal@uon.edu.au)
Newcastle University https://orcid.org/0000-0001-9728-8001
Shah Jahan Miah
The University of Newcastle Business School https://orcid.org/0000-0002-3783-8769

Research Article

Keywords: Machine learning, clinical, literature review, systematic review, Latent Dirichlet Allocation, topic modeling

Posted Date: January 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1090813/v1

Abstract

Big data analytics utilizes different analytics techniques to transform large volume and diversified big dataset. The analytics uses various computational methods such as different Machine Learning (ML) in convert raw data to valuable insights. The ML assist individuals to perform work activities quicker and better, and empower decision-makers in system use. Since academics and industry practitioners have growing interests on ML, how different applications of ML in specific problem domains have been explored, but not in a holistic manner from the past literature. This paper aims to promote the utilization of intelligent literature review for researchers by introducing a step-by-step framework on a case providing the code template. We offer an intelligent literature review to obtain in-depth analytical insight of ML applications in the clinical domain to: a) develop the intelligent literature framework using traditional literature and Latent Dirichlet Allocation (LDA) topic modeling, b) analyze research documents using traditional systematic literature review revealing ML applications, and c) identify topics from documents using LDA topic modeling. We used a PRISMA framework for the traditional literature review, reviewed four databases (e.g. IEEE, PubMed, Scopus, and Google Scholar), which are published between 2016 and 2021 (September). The framework comprises two stages – Traditional systematic literature review and LDA topic modeling. The intelligent literature review framework reviewed 305 research documents in a transparent, reliable, and faster way.

1. Introduction

Organizations are globally, harnessing the power of various big data using various machine learning techniques and utilizing them to reshape their business frameworks. Big data analytics techniques analyze vast amount of data, which are known as ‘Big Data’ to uncover hidden patterns, untold associations, anomalies, and other perceptions. Big Data allude to the enormous amount of data that cannot handle with traditional database management system. It is characterized by 5 V’s sometimes, which refers to volume, variety, velocity, veracity, and value [29]. Machine Learning (ML) is a one kind of big data analytics technique, which is a rapidly growing sub-field in information sciences that deals with numerous methods for machines to learn from past experiences (e.g. past datasets) without explicitly doing the traditional programming [2, 22]. Clinical care enterprises face a huge challenge due to the increasing utilization of technologies to improve clinical care outcomes, while costs hinder. For example, an electronic health record contains a huge amount of patient information, drug administration, imaging data using various modalities. The variety and quantity of data render the clinical domain an ideal topic to appraise the value of ML in research.

ML tools have become central focus to modern biomedical research because of better admittance to large datasets, exponential processing power, and key algorithmic developments allowing ML models to handle increasingly challenging data [28]. Different ML approaches can analyze a huge amount of data, including difficult and abnormal patterns. Most studies have focused on ML and its impacts on clinical practices [7, 20, 21, 22, 25, 26]. Fewer studies have examined the process of utilization of ML algorithms [23, 24, 27].

ML becomes an interdisciplinary science that integrates computer science, mathematics, and statistics. It is also a methodology that builds smart machines for artificial intelligence. Its applications comprise algorithms—an assortment of instructions to perform specific tasks—crafted to independently learn from data without human intercession. Over the time, ML algorithms improve their prediction accuracy without a need for programming. Based on this, we offer an intelligent literature review using traditional literature review and Latent Dirichlet Allocation (LDA) topic modeling in the clinical domain. Theoretical measures direct the current study results because previous literature provides a strong foundation for future IS researchers to investigate ML in the clinical sector. The main aim of this study is to develop the intelligent literature framework using traditional literature using four digital databases -IEEE, Google Scholar, PubMed, and Scopus then performed LDA topic modeling, which can assist healthcare or clinical researchers in analyzing many documents with little effort and a small amount of time.

2. Methodology

As traditional systematic literature is destined to be obsolete, which is time-consuming with restricted processing power, resulting in a lower number of sample documents investigated. Both academic and practitioner researchers frequently require to discover, organize, and comprehend new and unexplored research areas. As a part of a traditional literature review that involves an enormous number of papers, the choice for a researcher is either to restrict the number of papers to review a priori or analyze the review using some other methods. The proposed intelligent literature review framework assists future researchers in using appropriate technology, producing accurate results, and saving time. We present the framework below in figure 1.

We follow traditional systematic literature review methodologies [3, 6, 17], including a PRISMA framework [30]. We review four digital databases and develop three stages entailing planning, conducting, and reporting the review (Figure 2).

2.1 Planning the review

1. **Research articles —**The research articles are classified using some keywords as mentioned below in Table 1.
2. **Digital database**: Four databases (IEEE, PubMed, Scopus, Google Scholar) were used to collect details for reviewing research articles.
3. **Review protocol development**: We first used Scopus to search the information and found many studies regarding this review. We then searched PubMed, IEEE, and Google scholar for articles and extracted only relevant papers matching our keywords and review context based on their full-text availability.
4. **Review protocol evaluation**: To support the selection of research articles and inclusion and exclusion criteria, the quality of articles was explored and assessed to appraise their suitability and impartiality [13]. Only articles with keywords “machine learning” and “clinical” in document titles and abstracts were selected.

Table 1
Inclusion criteria

Inclusion Criteria	Description
1	Keywords (or short phrases) include: “Machine Learning,” “Machine Learning application,” “Machine Learning algorithms,” “Machine Learning techniques,” “Clinical,” “Clinical domain,” “Clinical sector.” Operators of search syntax are OR, AND. AND operator signifies that both keywords must be present in the search queries, and OR means that at least one keyword must be present in the queries searched
2	Research articles published between 2016 and 2021 (September)
3	Research articles published in English
4	Research limited to journal and conference articles
5	Only full-text articles

Table 2
Exclusion criteria

Exclusion Criteria	Description
1	Exclude duplicate research articles with matching title and/or digital object identifier (DOI)
2	Non-English research articles

2.2 Conducting the review

The second step is conducting the review, which includes a description of Search Syntax and data synthesis.

2.2.1 Search Syntax: Table 3 details the syntax used to select research articles

Table 3
Search Syntax for selected research articles

Database	Search syntax
Scopus	TITLEA ND ABSTRACT (machine learning in clinical) AND PUBYEAR > 2015 AND (LIMIT-TO (OA,"all")) AND (LIMIT-TO (LANGUAGE,"English")) AND (LIMIT-TO (PUBSTAGE,"final")) AND (LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"re") OR LIMIT-TO (DOCTYPE,"cp")).
IEEE	"Document Title and Abstract": machine learning in clinical
PubMed	(("machine"[All Fields] OR "machines"[All Fields]) AND "learning in clinical"[Title]) AND (2016:2021[pdat])
Google Scholar	allintitle: "machine learning in clinical"

2.2.2 Data Synthesis

We used a qualitative meta-synthesis technique to understand the methodology, algorithms, applications, qualities, results, and current research impediments. Qualitative meta-synthesis is a coherent approach for analyzing data across qualitative studies [1]. Our first search identified 534,327 papers, comprising Scopus (24,498), IEEE (2558), PubMed (11,271), and Google Scholar (496,000) articles with the selected keywords. After subjecting this dataset to our inclusion and exclusion criteria, articles were reduced to Scopus (181), IEEE (62), PubMed (37), and Google Scholar (46) (Figure 3).

2.3 Conversion of pdf files to a text document

The Python coding is used to convert pdf files which are shared on GitHub (https://github.com/MachineLearning-UON/Topic-modeling-using-LDA/blob/c11dfc94d552f6c08a4244a078d64583aa7e31b3/Conversion%20PDF%20to%20TXT%20file). The one text document is prepared with 305 research papers collected from a traditional literature review.

2.4 Topic Modelling for Intelligent literature review

Our Intelligent literature review is developed using a combination of traditional literature review and topic modeling [29]. We use topic modeling—probability generating, a text-mining technique widely used in computer science for text mining and data recovery. Topic modeling used in numerous papers to analyze [10, 12, 18] and use various ML algorithms [5] such as Latent Semantic Indexing (LSI), Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), Parallel Latent Dirichlet Allocation (PLDA), and Pachinko Allocation Model (PAM). We developed a new methodology framework using LDA as it is most widely and easily used [9, 10, 11] very elementary [2]. LDA is an unsupervised, probabilistic ML algorithm that discovers topics by calculating patterns of word co-occurrence across many documents or corpus. Each LDA topic is distributed across each document as a probability.

While there are numerous ways of conducting a systematic literature review, most strategies require a high expense of time and prior knowledge of the area in advance. This study examined the expense of various text categorization strategies, where the assumptions and cost of the strategies are analyzed [5]. Interestingly, except manually reading the papers and topic modeling, all the strategies require prior knowledge of the papers' categories and have a high pre-examination cost. However, topic modeling can be automated, alternate the utilization of researchers' time, demonstrating a perfect match for the utilization of topic modeling as a part of an Intelligent literature review. The utilization of topic modeling has been used in a few papers to categorize research papers presented in Table 4.

Table 4
Topic modeling applications

Author Details	Data type	Topic modeling used	Intended aim	Data size
Quinn et al. (2010) [5]	Legislative Speech	Own implemented method	To develop a statistical learning model	118,000 speeches (70,000,000 words)
DiMaggio et al. (2013) [10]	Newspapers	LDA	Identifying concepts in news coverage	8000
Koltsova & Koltcov (2013) [12]	Web posts	LDA	Explore the political agenda for live journal	1,300,000
Grimmer (2010)[4]	Press release	Own implemented method	To develop a model	24,000
Maier et al. (2018)[18]	Web documents	LDA	Explore the validity and reliability of the LDA model	186,557 web documents

The papers in the above table analyzed speeches, web documents, web posts, press releases, and newspapers. However, none of those papers have developed the framework to perform traditional literature reviews from digital databases and then use topic modeling to save time. However, this research points out the utilization of LDA in academics and explores four parameters – text pre-processing, model parameters selection, reliability, and validity [18]. Topic modeling identifies patterns of the repetitive word across a corpus of documents. Patterns of word co-occurrence are conceived of as hidden ‘topics’ which are available in the corpus. First, documents must be modified to be machine-readable, with only their most informative features used for topic modeling. We modify documents in a three-stage process entailing pre-processing, topic modeling, and post-processing.

The utilization of topic modeling presents an opportunity for researchers to use advanced technology for the literature review process. Topic modeling has been utilized online and requires many statistical skills, which not all researchers have. Therefore, we have shared the codes in GitHub with the default parameter for future researchers.

2.4.1 Pre-Processing

Székely and Brocke [15] explained that pre-processing is a seven-step process which explored below and mentioned in figure 1 as part B:

1. Load data
2. Optical character recognition
3. Filtering non-English words
4. Document tokenization
5. Text cleaning
6. Word lemmatization
7. Stop word removal

2.4.2 Topic Modelling using LDA

Several research articles have been selected to run LDA topic modeling which explained in Table 4. LDA model results present the coherence score for all the selected topics and a list of the most frequently used words for each.

2.4.3 Post-Processing

The goal of the post-processing stage is to identify and label topics and topics relevant for use in the literature review. The result of the LDA model is presented as a list of topics and probabilities of each document (paper). The list is utilized to assign a paper to a topic by arranging the list by the highest probability for each paper for each topic. All the topics contain papers that are like each other. To reduce the risk of error in topic identification, a combination of inspecting the most frequent words for each topic and a paper view is used. After the topic review, it will present in the literature review.

Following the intelligent literature review, results of the LDA model should be approved or validated by statistical, semantic, or predictive means. Statistical validation defines the mutual information tests of result fit to model assumptions; semantics validation requires hand-coding to decide if the importance of specific words varies significantly and as expected with tasks to different topics which is used in the current study to validate LDA model result; and predictive validation refers to checking if events that ought to have expanded the prevalence of particular topic if out interpretations are right, did so [8, 11].

3. Results

Our systematic literature review identified 305 research papers after performing a traditional literature review. After executing LDA topic modeling, only 115 articles show the relevancy with our topic "machine learning application in clinical domain'. The following stages present LDA topic modeling process.

3.1 Pre-processing

The 305 research papers were stacked into a Python environment, then converted into a single text file. The seven steps have been carried out, which were described earlier in 2.4.1.

3.2 Topic modeling

The two main parameters of the LDA topic model are the dictionary (id2word)- dictionary and the corpus – doc_term_matrix. The LDA model is created by running the command:

Creating the object for LDA model using gensim library

LDA = gensim.models.ldamodel.LdaModel

Build LDA model

lda_model = LDA(corpus=doc_term_matrix, id2word=dictionary, num_topics=20, random_state=100,

chunksize=1000, passes=50,iterations=100)

In this model, ‘num_topics’ = 20, ‘chunksize’ is the number of documents used in each training chunk, and ‘passes’ is the total number of training passes.

Firstly, the LDA model is built with 20 topics; each topic is represented by a combination of 20 keywords, with each keyword contributing a certain weight to a topic. Topics are viewed and interpreted in the LDA model, such as Topic 0, represented as below:

(0,

'0.005*"analysis" + 0.005*"study" + 0.005*"models" + 0.004*"prediction" + 0.003*"disease" + 0.003*"performance" + 0.003*"different" + 0.003*"results" + 0.003*"patient" + 0.002*"feature" + 0.002*"system" + 0.002*"accuracy" + 0.002*"diagnosis" + 0.002*"classification" + 0.002*"studies" + 0.002*"medicine" + 0.002*"value" + 0.002*"approach" + 0.002*"variables" + 0.002*"review"'),

Our approach to finding the ideal number of topics is to construct LDA models with different numbers of topics as K and select the model with the highest coherence value. Selecting the ‘K' value that denotes the end of the rapid growth of topic coherence ordinarily offers significant and interpretable topics. Picking a considerably higher value can provide more granular sub-topics if the 'K’ selection is too large, which can cause the repetition of keywords in multiple topics.

Model perplexity and topic coherence values are -8.855378536321144 and 0.3724024189689453, respectively. To measure the efficiency of the LDA model, is lower the perplexity, the better the model is. Topics and associated keywords were then examined in an interactive chart using the pyLDAvis package, which presents the topics are 20 and most salient terms in those 20 topics, but these 20 topics overlap each other, which means the keywords are repeated in these 20 topics and topics are overlapped, which means so decided to use num_topics = 9 and presented PyLDAvis Figure below. Each bubble on the left-hand side plot represents a topic. The bigger the bubble is, the more predominant that topic is. A decent

topic will have a genuinely big, non-overlapping bubble dispersed throughout the graph instead of grouped in one quadrant. A topic model with many topics will typically have many overlaps, small-sized bubbles clustered in one locale of the graph, as shown in Figure 4.

One of the practical applications of topic modeling is discovering the topic in a provided document. We discover the topic number with the highest percentage contribution in that document, as shown in Figure 5.

3.3 Post-processing

The next stage is to process the findings after performing LDA topic modeling. The topic name is displayed with the topic number from 0 to 8, which represents in the below table, which includes the Topic number and Topic words.

Table 5
The number of topics and topic words in the LDA result

Topic number	Topic words
0	analysis, study, models, prediction, disease, performance, different, results, patient, feature
1	models, prediction, study, analysis, disease, results, patient, performance, studies, accuracy
2	study, models, prediction, analysis, results, disease, performance, feature, neural, accuracy
3	models, study, analysis, patient, feature, prediction, system, studies, results, training
4	study, models, analysis, prediction, patient, disease, information, accuracy, training, results
5	prediction, analysis, models, study, disease, patient, information, results, validation, training
6	study, analysis, feature, models, performance, prediction, disease, studies, patient, accuracy
7	study, models, disease, patient, prediction, results, performance, analysis, accuracy, algorithms
8	study, disease, analysis, studies, models, accuracy, prediction, performance, decision, algorithm

The result represents the percentage of the topics in all documents, which presents that topic 0 and topic 6 have the highest percentage and used in 58 and 57 documents, respectively, with 115 papers. The result of this research was an overview of the exploration areas inside the paper corpus, addressed by 9 topics.

4. Discussion

The study has introduced an intelligent literature review framework that uses ML to analyze existing research documents or articles. We demonstrate how topic modeling can assist literature review by reducing the manual screening of huge quantities of literature for more efficient use of researcher time. An LDA algorithm provides default parameters and data cleaning steps, reducing the effort required to review literature. An additional advantage of our framework is that the n literature review provides accurate results with little time, and it comprises traditional ways to analyze literature and LDA Topic modeling.

This framework is constructed in a step-by-step manner. It can be used efficiently by researchers because it requires less technical knowledge than other ML algorithms. There is no restriction on the quantity of the research papers that it can measure. This research extends knowledge to similar studies in this field [14, 16, 19, 22, 29], which present topic modeling. The study acknowledges the inspiring concept of smart literature defined by Asmussen, & Møller [31]. The researchers previously provided a brief description of how LDA utilised in topic modelling. Our research followed the basic idea but enhance its significance to broaden its scale and focusing on specific domain such as clinical domain for producing insights from existing research articles. Our research developed the intelligent framework on that, which is combination of traditional literature review and topic modelling using LDA which provide more accurate and transparent results. The results are shared via public access on GitHub using this link https://github.com/MachineLearning-UON/Topic-modeling-using-LDA/blob/c11dfc94d552f6c08a4244a078d64583aa7e31b3/LDA%20Topic%20modeling.

5. Conclusion

A framework that empowers researchers to use topic modeling for rapidly and reliably investigating a limitless number of papers, reducing their need to read each individually, is developed. Topic modeling using the LDA algorithm can assist future researchers as they often need an outline of various research fields with minimal pre-existing knowledge. The proposed framework can empower researchers to review more papers in less time with more accuracy. Our intelligent literature review framework includes a holistic literature review process (conducting, planning, and reporting the review) and an LDA topic modeling (pre-processing, topic modeling, and post-processing stages), which conclude the results of 115 research articles are relevant to the search. The automation of topic modeling with default parameters could also be explored to benefit non-technical researchers to explore topics or related keywords in any problem domain.

Abbreviations

IEEE—The Institute of Electrical and Electronics Engineers

ML—Machine Learning

LDA—Latent Dirichlet Allocation

OC—Organizational Capacity

LSI—Latent Semantic Indexing

LSA—Latent Semantic Analysis

NPF—Non-Negative Matrix Factorization

PLDA—Parallel Latent Dirichlet Allocation

PAM—Pachinko Allocation Model

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Data will be supplied upon request

Competing interests

Not applicable

Funding

Not applicable

Authors' contributions

The first author conducted the research, while the second author has ensured quality standards and rewritten the entire findings linking to underlying theories.

Acknowledgments

Not applicable

References

1. Beck CT. A meta-synthesis of qualitative research. MCN: The American Journal of Maternal/Child Nursing. 2002;27(4):214-21.

2. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. the Journal of machine Learning research. 2003;3:993-1022.

3. Rowley J, Slack F. Conducting a literature review. Management research news. 2004.

4. Grimmer J. A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. Political Analysis. 2010;18(1):1-35.

5. Quinn KM, Monroe BL, Colaresi M, Crespin MH, Radev DR. How to analyze political attention with minimal assumptions and costs. American Journal of Political Science. 2010;54(1):209-28.

6. Rozas LW, Klein WC. The value and purpose of the traditional qualitative literature review. Journal of evidence-based social work. 2010;7(5):387-99.

7. Mimno D, Blei D, editors. Bayesian checking for topic models. Proceedings of the 2011 conference on empirical methods in natural language processing; 2011.

8. Blei DM. Probabilistic topic models. Communications of the ACM. 2012;55(4):77-84.

9. DiMaggio P, Nag M, Blei D. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding. Poetics. 2013;41(6):570-606.

10. Grimmer J, Stewart BM. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political analysis. 2013;21(3):267-97.

11. Koltsova O, Koltcov S. Mapping the public agenda with topic modeling: The case of the Russian livejournal. Policy & Internet. 2013;5(2):207-27.

12. Ouhbi S, Idri A, Fernández-Alemán JL, Toval A. Requirements engineering education: a systematic mapping study. Requirements Engineering. 2015;20(2):119-38.

13. Greene D, Cross JP. Exploring the political agenda of the european parliament using a dynamic topic modeling approach. Political Analysis. 2017;25(1):77-94.

14. Székely N, Vom Brocke J. What can we learn from corporate sustainability reporting? Deriving propositions for research and practice from over 9,500 corporate sustainability reports published between 1999 and 2015 using topic modelling technique. PloS one. 2017;12(4):e0174807.

15. Abuhay TM, Kovalchuk SV, Bochenina K, Mbogo G-K, Visheratin AA, Kampis G, et al. Analysis of publication activity of computational science society in 2001–2017 using topic modelling and graph theory. Journal of computational science. 2018;26:193-204.

16. Li S, Wang H. Traditional literature review and research synthesis. The Palgrave handbook of applied linguistics research methodology. 2018:123-44.

17. Maier D, Waldherr A, Miltner P, Wiedemann G, Niekler A, Keinert A, et al. Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. Communication Methods and Measures. 2018;12(2-3):93-118.

18. Behera RK, Bala PK, Dhir A. The emerging role of cognitive computing in healthcare: a systematic literature review. International journal of medical informatics. 2019;129:154-66.

19. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. Genome medicine. 2019;11(1):1-12.

20. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating AI and practical implications. Yearbook of medical informatics. 2019;28(01):128-34.

21. Shah P, Kendall F, Khozin S, Goosen R, Hu J, Laramie J, et al. Artificial intelligence and machine learning in clinical development: a translational perspective. NPJ digital medicine. 2019;2(1):1-5.

22. Mårtensson G, Ferreira D, Granberg T, Cavallin L, Oppedal K, Padovani A, et al. The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. Medical Image Analysis. 2020;66:101714.

23. Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. JMIR medical informatics. 2020;8(3):e17984.

24. Weng W-H. Machine learning for clinical predictive analytics. Leveraging Data Science for Global Health: Springer, Cham; 2020. p. 199-217.

25. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. Med. 2021.

26. Chang C-H, Lin C-H, Lane H-Y. Machine Learning and Novel Biomarkers for the Diagnosis of Alzheimer's Disease. International Journal of Molecular Sciences. 2021;22(5):2761.

27. Connor KL, O'Sullivan ED, Marson LP, Wigmore SJ, Harrison EM. The Future Role of Machine Learning in Clinical Transplantation. Transplantation. 2021;105(4):723-35.

28. Hassan N, Slight R, Weiand D, Vellinga A, Morgan G, Aboushareb F, et al. Preventing sepsis; how can artificial intelligence inform the clinical decision-making process? A systematic review. International Journal of Medical Informatics. 2021:104457.

29. Kushwaha AK, Kar AK, Dwivedi YK. Applications of big data in emerging management disciplines: A literature review using text mining. International Journal of Information Management Data Insights.

2021;1(2):100017.

30. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Bmj. 2021;372.

Figures

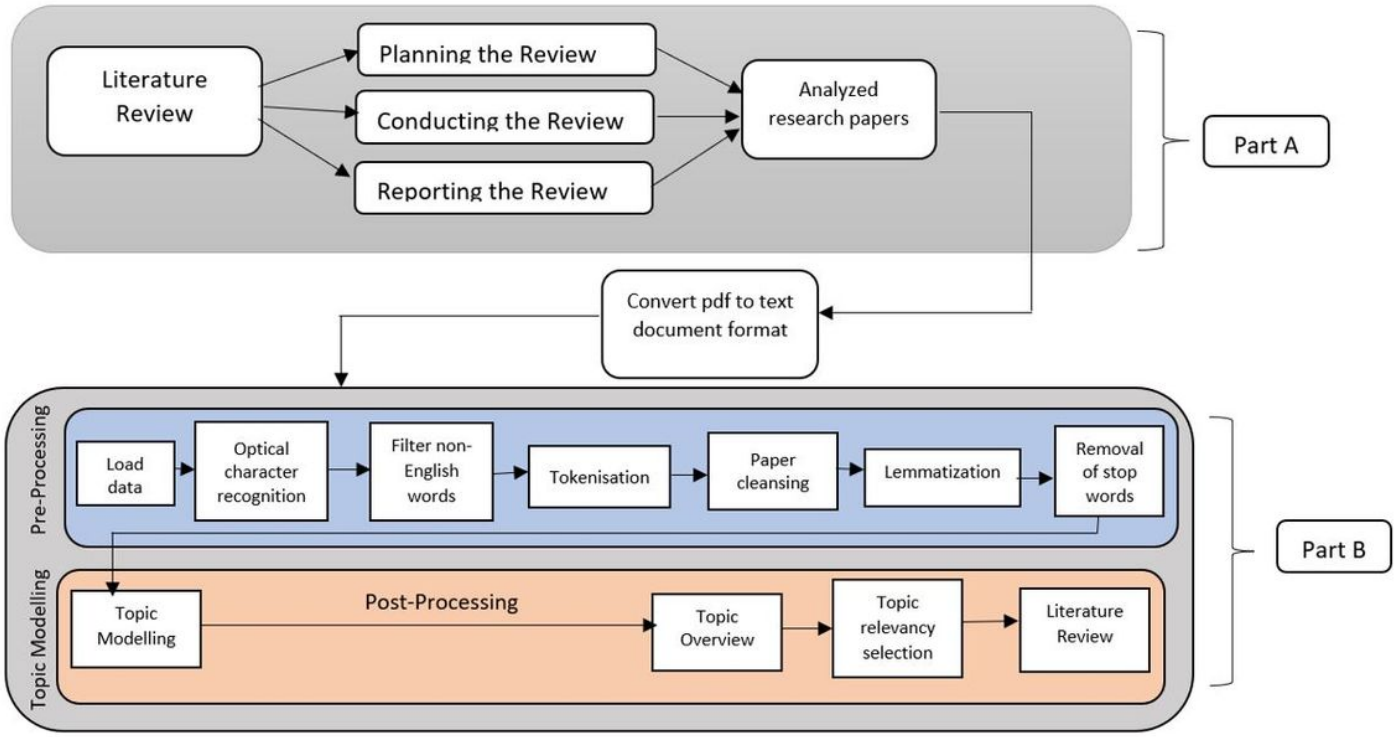

Figure 1

Proposed intelligent literature review framework

Figure 2

Traditional Literature review three stages

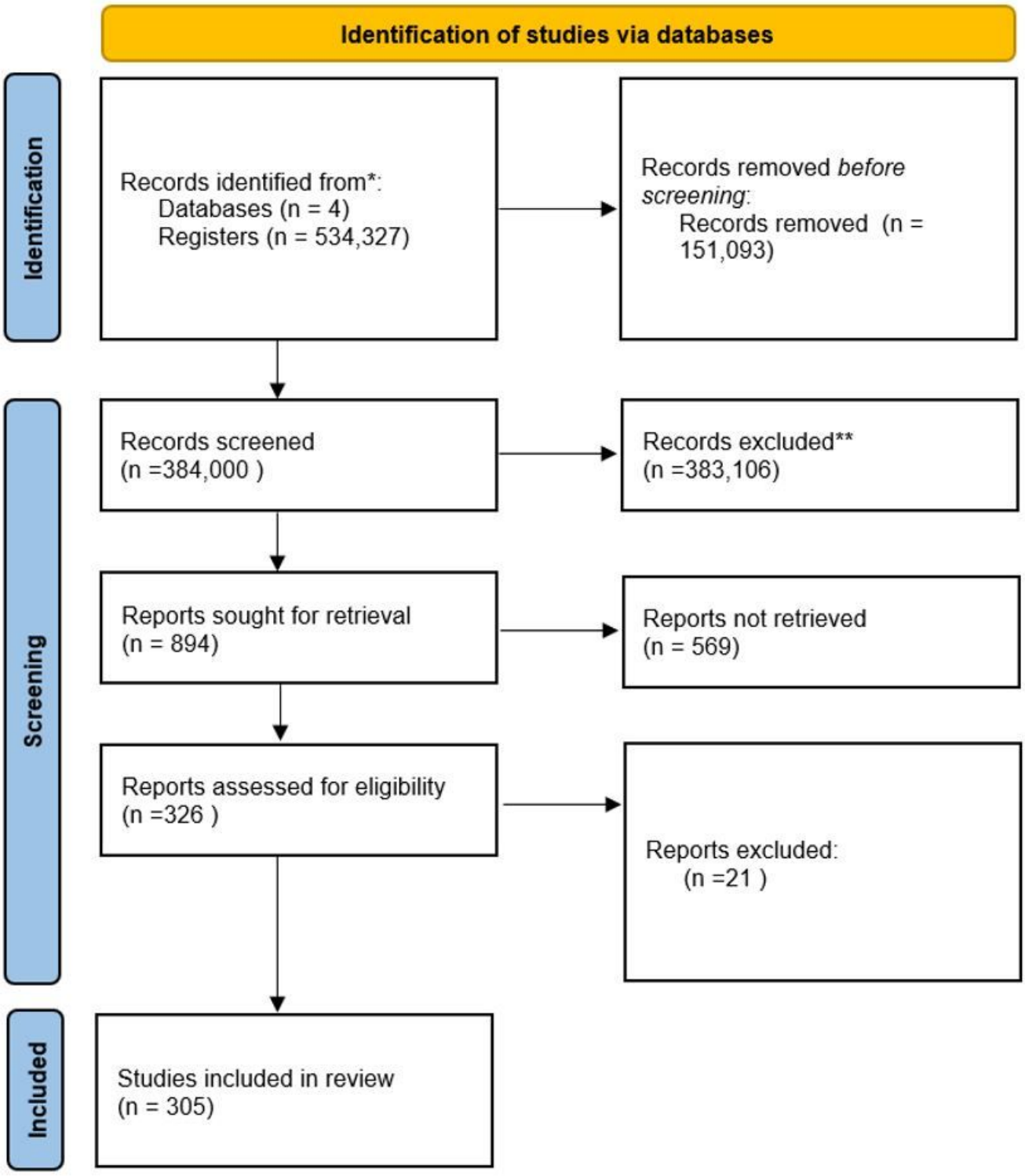

Figure 3

PRISMA framework of traditional literature review

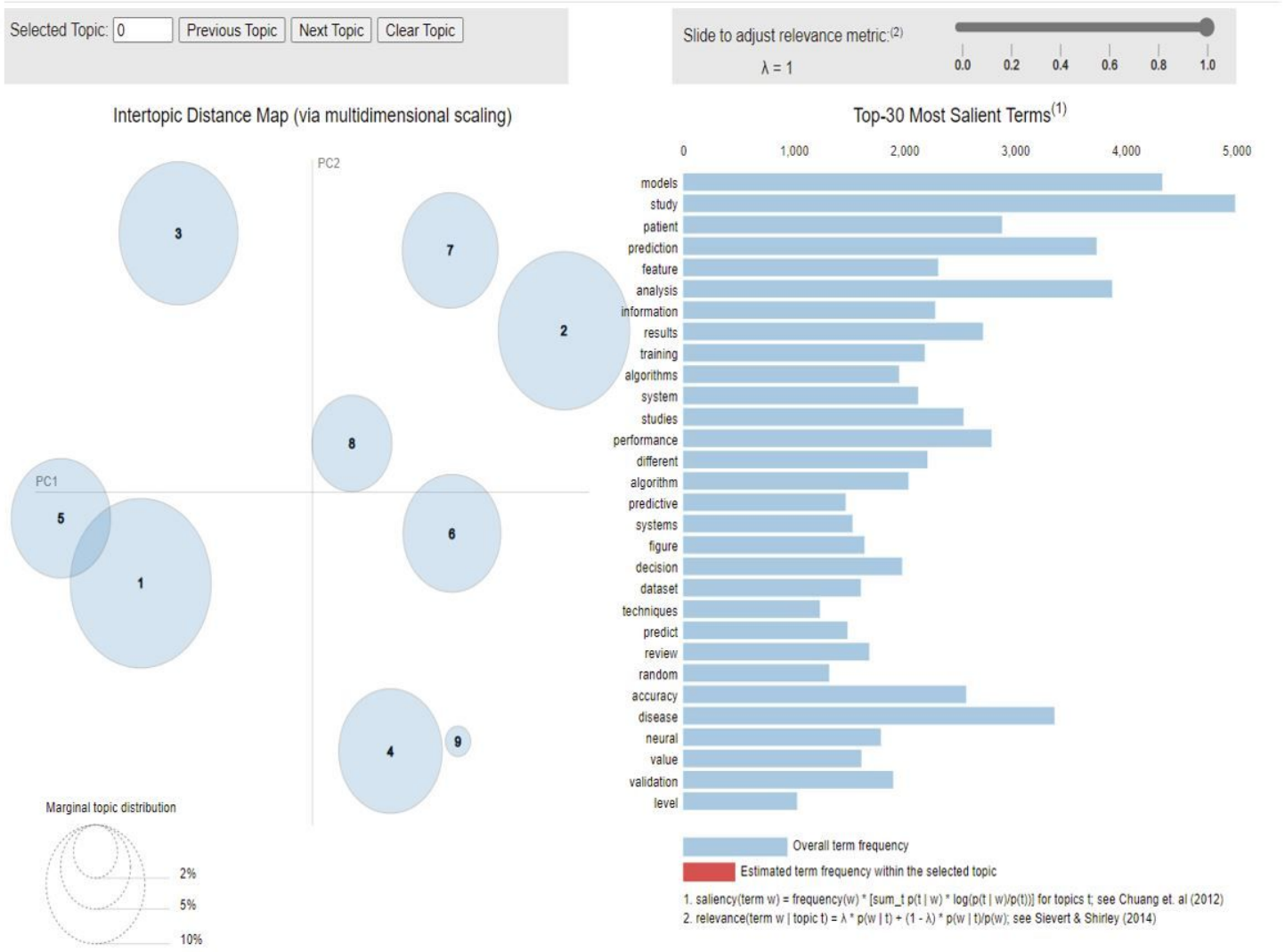

Figure 4

PyLDAvis graph with nine vital topics in clinical domain

	Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	0	8.0	0.5048	study, disease, analysis, studies, models, acc...	[th, international, conference, advanced, tech...
1	1	6.0	0.6078	study, analysis, feature, models, performance,...	[comparative, study, techniques, systems, cont...
2	2	6.0	0.9014	study, analysis, feature, models, performance,...	[international, conference, computational, sci...
3	3	1.0	0.5276	models, prediction, study, analysis, disease, ...	[article, comparison, models, versus, evaluati...
4	4	2.0	0.9884	study, models, prediction, analysis, results, ...	[international, conference, big, knowledge, ic...
5	5	8.0	0.6417	study, disease, analysis, studies, models, acc...	[authorized, licensed, limited, newcastle, dow...
6	6	0.0	0.5757	analysis, study, models, prediction, disease, ...	[open, amendments, paper, approach, forecast, ...
7	7	3.0	0.4755	models, study, analysis, patient, feature, pre...	[article, approach, predict, extreme, inactivi...
8	8	5.0	0.7622	prediction, analysis, models, study, disease, ...	[international, conference, big, big, article,...
9	9	0.0	0.6624	analysis, study, models, prediction, disease, ...	[original, published, march, fonc, radiomics, ...
10	10	0.0	0.6840	analysis, study, models, prediction, disease, ...	[algorithm, assessment, metabolomic, fingerpri...
11	11	2.0	0.6315	study, models, prediction, analysis, results, ...	[review, article, page, narrative, review, pro...
12	12	4.0	0.4534	study, models, analysis, prediction, patient, ...	[received, september, accepted, october, date,...
13	13	6.0	0.8737	study, analysis, feature, models, performance,...	[contents, lists, available, sciencedirect, bi...
14	14	4.0	0.5580	study, models, analysis, prediction, patient, ...	[computers, biology, medicine, contents, lists...
15	15	0.0	0.8048	analysis, study, models, prediction, disease, ...	[salem, bmc, inform, decis, mak, open, access,...
16	16	1.0	0.8108	models, prediction, study, analysis, disease, ...	[connolly, bmc, bioinformatics, open, access, ...
17	17	6.0	0.8837	study, analysis, feature, models, performance,...	[contents, lists, available, sciencedirect, bi...
18	18	0.0	0.7735	analysis, study, models, prediction, disease, ...	[salem, bmc, inform, decis, mak, open, access,...
19	19	1.0	0.7828	models, prediction, study, analysis, disease, ...	[contents, lists, available, sciencedirect, ps...

Figure 5

Dominant topics with topic percentage contribution