MR-v1

by M R

Submission date: 06-Jan-2023 10:28PM (UTC+0500)
Submission ID: 1989272612
File name: EVIDENCE_RETRIEVAL_FOR_EBM_CONTENT.docx (53.17K)
Word count: 1902
Character count: 11312

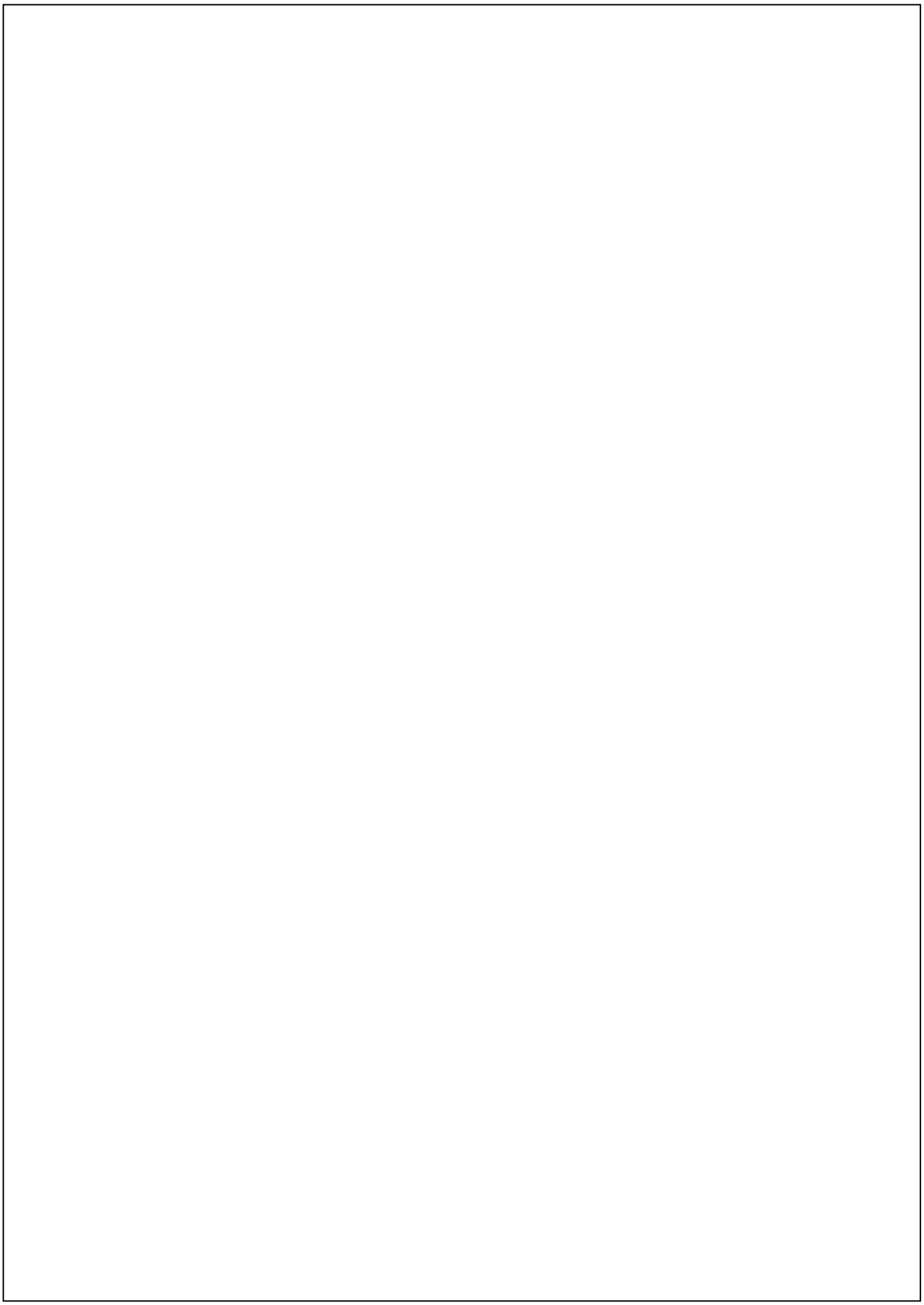
Abstract

Evidence-Based Medicine (EBM) has become a decision support system for medical consultants. There is a vast amount of published literature in the medical field, and searching for the most relevant literature is challenging for those who do have not the necessary expertise. Finding the right literature requires specialized knowledge and skills to identify and evaluate the quality and relevance of the information.

The purpose of an EBM-based system is, to retrieve for the Medical Practitioners (usually Physicians), evidence-based latest relevant information on disease and cure, typically based on the keywords extracted from the text that carries the patient's details and medical records of the patients. These systems widely use a Patient Intervention Comparison and Outcome (PICO) model to formulate a query that helps classify and identify the patient's diseases.

A large amount of information is available on different diseases on the internet. In this work, we will reduce the gap between the formulated query and the retrieved documents or information by improving the relevancy and accuracy of the retrieved information by means of extracting and ranking the evidence available for the relevant information.

Keywords: [Information Retrieval, Multi-Class Classification, Ranking, Machine Learning. Evidence Retrieval]



INTRODUCTION

EBM is "A process of turning clinical problems into questions and then systematically locating, appraising, and using contemporaneous research findings as the basis for clinical decisions". (Rosenberg & Donald,1995)

EBM involves using effective information search strategies to find reliable, current information from various sources, and using extraction strategies to efficiently collect and analyze the retrieved information. This process includes a step called the Critical Appraisal Exercise (CAE).(Sarker et al., 2017). To achieve the Critical Appraisal Exercise we have to follow the following steps according to (Selvaraj et al., 2010):

1. Identifying the specific issues or problems that a patient is experiencing, and determining what evidence is necessary to address those issues.
2. Search the literature review.
3. Choosing the most reliable studies and using EBM guidelines to assess their validity.
4. Evaluating the quality of the evidence.
5. Extracting and synthesizing relevant evidence, and using it to address the current problem or issue

To comply with CAE practice for achieving EBM, practitioners must face multiple challenges like i) Crafting an appropriate query ii) Analyzing the retrieved literature to find the most relevant and up to date information and iii) Evaluating the evidence of the literature.

To overcome these challenges, usually, a PICO-based syntax is used in Evidence-Based Medicine to Retrieve literature. Moreover, quality assessment criteria are required to retrieve good quality research.

Classical information retrieval models may not be effective in certain domains, such as Evidence-based Medicine, because they rely on simply identifying the presence or absence of certain terms. To better retrieve and rank relevant documents in this complex domain, an additional strategy that considers the quality and relevance of the evidence and methodology must be employed in conjunction with classical information retrieval models.(Serrano-Guerrero at el., 2020)

To search the relevant literature, a query is formulated to retrieve results. In this work, we are focusing on the third and fourth processes of CAE. We have designed quality criteria by assessing the relevancy of the retrieved literature. Based on these Quality Criteria, we have proposed an **Evidence Retrieval System using the Patient Intervention Comparison and Outcome (PICO) framework**.

The composition of the work is as follows: a) Brief overview of this work is presented in this section. b) The relevant recent literature explains the current research in EBM following the introduction. c) The methodology describes our proposed approach to contribute to EBM and specifically in pandemic literature. d) The experiments, results, and discussions are yet to be provided

LITERATURE REVIEW

The literature review in this study aims to provide a comprehensive overview of the current state of knowledge on the topic of evidence-based medicine (EBM) in healthcare. EBM is an approach to healthcare that involves using the best available evidence from research to inform clinical decision-making, with the goal of improving patient outcomes. In this literature review, we aim to identify and synthesize the existing research on EBM, highlighting key findings and identifying gaps in the literature. This review will provide a foundation for the research presented in this paper and will help to contextualize the study within the broader field of EBM research.

(Esteva et al.,2021) focus to create a search engine called CO-search that could handle complex queries related to COVID literature. In the medical field, even small changes in wording or the order of words can significantly alter the meaning of a query, making it challenging to accurately search for information. To address this issue, the study employed a combination of algorithms, including Siamese-BERT for query-level encoding, BM25 and TF-IDF for determining the importance of words in a query, and question answering and abstractive summarization for re-ranking documents. While the study was successful in handling many complex queries, it identified areas where improvement is needed, particularly in terms of its keyword model. One potential next step for this research could be to apply the same approach to other diseases beyond COVID. The dataset used in this study is CORD-19 (Lu Wang et al.,2020) which has 41000 publications in its initial release.

Another work related to COVID has been done by (Voorhees et al., 2020) to assess the effectiveness of search algorithms and systems in aiding individuals such as scientists, clinicians, policymakers, and others in navigating and utilizing the vast amount of literature on COVID-19 and Exploring approaches for effectively organizing and accessing scientific information during future global health emergencies.

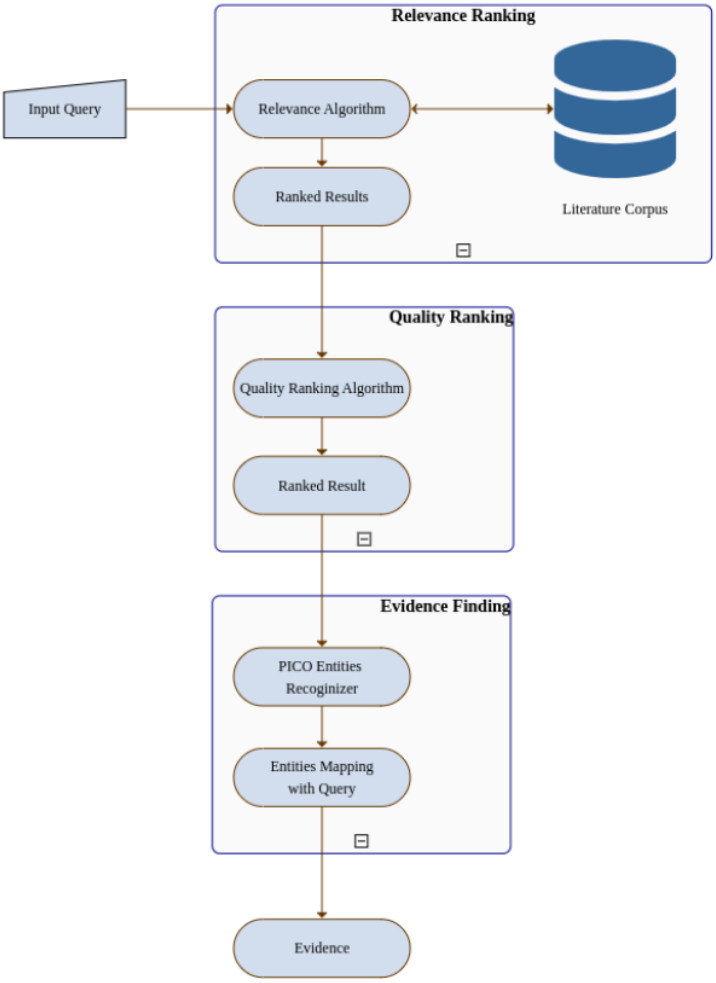
(Stylianou et al., 2020) has contributed to the EBM by finding evidence using the PICO Entity recognizer and PICO statement classifier. Their proposed framework comprises two steps. Recognize PICO entities by using seq-seq custom NER model and also 2D CNN, ELMo, and Bi-LSTM for capturing character information, embeddings, and vector representation respectively, and, classify the PICO sentences that answer the clinical questions by using Multinomial Naive Bayes, Linear SVC, Gradient Boosting, Extreme Gradient Boosting, Decision Trees and Random Forest as classification models and word2vec for word representation and average pooling for sentence representation.

EBM is a specialized domain so, classical information retrieval models may struggle to effectively search for documents in this complex domain, where specific strategies are necessary to retrieve and rank documents based on their quality of evidence and methodology. These models can only consider the presence or absence of terms, making it necessary to supplement them with additional techniques to accurately search for relevant documents in this domain (Serrano-Guerrero et al., 2020). To overcome this problem, they designed their own algorithm. The study works as an extension of (Choi et al.,2012) and (Demner-Fushman& Lin, 2007) work. The study applies two-step ranking: 1)From query retrieved literature and ranked them according to the relevant criteria and form clusters. 2)Rank by quality criteria and return the result of the best-matching cluster. They also mention the future directions of adding feedback systems for re-ranking.

(Stylianou & Vlahavas, 2021) works in finding evidence by applying argument mining and finding their relations in form of graphs by using transformers and BERT for sentence embedding and Transformer decoder for argument relation classification. They claim to enable fine-grained searches of available medical literature using PICO elements as search parameters.

PROPOSED METHODOLOGY

The study proposed a system by taking ideas from (Stylianou et al., 2020), (Esteva et al., 2021), (Stylianou & Vlahavas., 2021) and, (Serrano-Guerrero et al., 2020). Our system named **Evidence Retrieval System using the Patient Intervention Comparison and Outcome (PICO) framework** consists of four major parts: 1) Retrieving relevant research articles. 2) Redeem high-quality literature by applying some quality measures. 3) Find the evidence from the literature and 4) Retrieve the result based on evidence.



Relevance Ranking

The first part of our system consists of two major components: the literature corpus and document relevance algorithm. The literature corpus includes literature on diseases. Each piece of literature is annotated as per the requirement of Information Retrieval (IR) algorithms.

The document relevancy algorithm will identify the relevance of each document to the input query. This assessment takes into account the complexities of the domain in determining the relevancy of the document.

Quality Ranking

After being processed by the relevance ranking algorithm, the input query is used as input for the quality ranking part. In this step, the quality ranking algorithm ranks the literature based on various attributes, such as the author, total citations, and publication year of the article.

Evidence Finding

With the ranking of the relevant and quality literature complete, the next phase of our system is to identify relevant evidence. To do this, we will first extract the PICO elements from the literature. PICO is a widely accepted format and this approach has been successfully used in many studies (Stylianou et al., 2020), (Stylianou & Vlahavas., 2021). After finding the PICO elements, we find relations of these elements with the input query.

Dataset

The CORD-19 corpus dataset (Lu Wang et al., 2020) is a collection of over 400,000 scientific papers about COVID-19, SARS-CoV-2, and earlier coronaviruses that was compiled by the U.S. White House, the National Institutes of Health, the Allen Institute for AI, the Chan-Zuckerberg Initiative, Microsoft Research, and Georgetown University in response to the global COVID-19 crisis. The dataset includes topics and relevance judgments made by experts in fields such as biology and medicine, which categorize topic-document pairs as irrelevant, partially relevant, or relevant. As of February 2021, the CORD-19 corpus is the largest publicly available collection of coronavirus literature. (Esteva et al., 2021)

References

1. Rosenberg, W., & Donald, A. (1995). Evidence based medicine: An Approach to clinical problem-solving. *BMJ*. https://doi.org/10.1136/bmj.310.6987.1122
2. Sarker, A., Molla, D., & Paris, C. (2017). Automated text summarisation and evidence-based medicine: A survey of two domains. https://doi.org/10.48550/arxiv.1706.08162
3. Stylianou, N., Razis, G., Goulis, D. G., & Vlahavas, I. (2020). EBM+: Advancing Evidence-Based Medicine via two level automatic identification of Populations, Interventions, Outcomes in medical literature. Artificial Intelligence in Medicine. https://doi.org/10.1016/j.artmed.2020.101949
4. Selvaraj, S., Yeshwant Kumar, N. N. T., Elakiya, M., Prarthana Saraswathi, C., Balaji, D., Nagamani, P., & Mohan, S. K. (2010). Evidence-based medicine - A new approach to teach medicine: A basic review for beginners. In Biology and Medicine.
5. Esteva, A., Kale, A., Paulus, R., Hashimoto, K., Yin, W., Radev, D., & Socher, R. (2021). COVID-19 information retrieval with deep-learning based semantic search, question answering, and abstractive summarization. *Npj Digital Medicine*, *4*(1). https://doi.org/10.1038/s41746-021-00437-0
6. Stylianou, N., & Vlahavas, I. (2021). TransforMED: End-to-End Transformers for Evidence-Based Medicine and Argument Mining in medical literature. *Journal of Biomedical Informatics*. https://doi.org/10.1016/j.jbi.2021.103767
7. Serrano-Guerrero, J., Romero, F. P., & Olivas, J. A. (2020). A relevance and quality-based ranking algorithm applied to evidence-based medicine. *Computer Methods and Programs in Biomedicine*. https://doi.org/10.1016/j.cmpb.2020.105415
8. Lu Wang, L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., Funk, K., Kinney, R., Liu, Z., Merrill, W., Mooney, P., Murdick, D., Rishi, D., Sheehan, J., Shen, Z.,

Stilson, B., Wade, A. D., Wang, K., Wilhelm, C., … Kohlmeier, S. (2020). CORD-19: The Covid-19 Open Research Dataset. ArXiv.

9. Choi, S., Ryu, B., Yoo, S., & Choi, J. (2012). Combining relevancy and methodological quality into a single ranking for evidence-based medicine. *Information Sciences*. https://doi.org/10.1016/j.ins.2012.05.027
10. Demner-Fushman, D., & Lin, J. (2007). Answering clinical questions with knowledge-based and statistical techniques. *Computational Linguistics*. https://doi.org/10.1162/coli.2007.33.1.63
11. Voorhees, E., Alam, T., Bedrick, S., Demner-Fushman, D., Hersh, W. R., Lo, K., Roberts, K., Soboroff, I., & Wang, L. L. (2020). TREC-COVID. *ACM SIGIR Forum*. https://doi.org/10.1145/3451964.3451965

MR-v1

ORIGINALITY REPORT

13%	10%	8%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	web.archive.org Internet Source	3%
2	Submitted to Arizona State University Student Paper	2%
3	www.ncbi.nlm.nih.gov Internet Source	2%
4	covid19dataportal.es Internet Source	1%
5	ikee.lib.auth.gr Internet Source	1%
6	Submitted to Liverpool John Moores University Student Paper	1%
7	Nikolaos Stylianou, Ioannis Vlahavas. "TransforMED: End-to-End Transformers for Evidence-Based Medicine and Argument Mining in medical literature", Journal of Biomedical Informatics, 2021 Publication	1%

8	www.world-today-news.com Internet Source	1 %
9	Jesus Serrano-Guerrero, Francisco P. Romero, Jose A. Olivas. "A relevance and quality-based ranking algorithm applied to evidence-based medicine", Computer Methods and Programs in Biomedicine, 2020 Publication	1 %
10	aclanthology.org Internet Source	1 %
11	"Trends of Artificial Intelligence and Big Data for E-Health", Springer Science and Business Media LLC, 2022 Publication	1 %
12	www.biorxiv.org Internet Source	1 %
13	Nikolaos Stylianou, Gerasimos Razis, Dimitrios G. Goulis, Ioannis Vlahavas. "EBM+: Advancing Evidence-Based Medicine via two level automatic identification of Populations, Interventions, Outcomes in medical literature", Artificial Intelligence in Medicine, 2020 Publication	<1 %
14	Sungbin Choi, Borim Ryu, Sooyoung Yoo, Jinwook Choi. "Combining relevancy and methodological quality into a single ranking	<1 %

for evidence-based medicine", Information Sciences, 2012
Publication

Exclude quotes	Off	Exclude matches	< 3 words
Exclude bibliography	On		