Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Design and Development of Efficient Electronic Medical Data Search Engine with Data Privacy Using Blockchain Technology

M.Karthika Devi[1], M.Thangaraj[2]

[1]Research Scholar Department of Computer Science at the School of Information Technology Madurai Kamaraj University, Madurai, India, Email:karthikamku2020@gmail.com, ORCID ID: 0009-0003-2501-3721.

[2]Professor and Head of the Department Department of Computer Science at the School of Information Technology Madurai Kamaraj University, Madurai, India, Email: thangarajmku@yahoo.com, ORCID ID :0000-0003-1296-8572

Abstract

Information retrieval from the medical repositories within a time frame is challenging task due to the rapid growth of patient's population. To address this challenge, a Gestalt Projection Tversky Censored Regressive Miyaguchi-Preneel cryptographic Blockchain (GPTCRMCB) is introduced. This GPTCRMCB method performs query preprocessing which includes tokenization and stop words removal. The targeted projection pursuit technique is used for extracting the keywords. Search engine calculates a relevance score by applying Tversky indexivecensored regression. Depend on similarity score value, relevant medical information is identifiedthroughsuperior accuracy. Finally, Miyaguchi-Preneel cryptographic hash Blockchain is employed for IR in a secure manner. Outcomes of GPTCRMCB achievedmaximum precision, recall, F1-score with minimal time than existing methods.

Keywords:Information Retrieval (IR), Electronic Medical Records(EMR) Search Engine, Gensim tokenizer, gestalt pattern matching, targeted projection pursuit technique, Tversky index, Blockchain

1. Introduction

Patient IR actsfundamentalpart in healthcare systems.EMR represent a majorimprovement in healthcare technology, offering transformative benefits in patient care, clinical decision-making and healthcare system efficiency. With the extensive exponential growth of medical data, the need for effective information retrieval methods has become important. Patient information retrieval plays a vital role in healthcare systems, facilitating access to relevant medical data for healthcare professionals. Efficient retrieval of patient information enables healthcare providers to make informed decisions, enhance patient care, and reorganize medical processes. EMRs database store a vast amount of structured and unstructured patient information, including medical history, diagnoses, lab results, imaging reports, and treatment plans. Retrieving relevant information from this database requires advanced techniques for healthcare information retrieval.

An Embedding-based knowledge structure (EKS) was designed [1] formedical IR. But it failed to improve accuracy. Aclustering-based method was introduced in [2]for increasing precision and recall. However, it failed to minimize time. Exponential Aquila Optimizer (EAO)-based Deep Fuzzy Clustering [3] was employed. But, it faced challenges for larger datasets. The blockchain system was presented [4] to indicate access controls. Study focused on health IR scheme [5] with fuzzy basis of ML. In [6], mixture dynamic filtering investigated. Lightweight safe information distribution examined [7]. However, it failed toprecision.

At [8], R-treeclustering performed. Study considered on Ontology-basis of plan [9]. During IR, EMR Search Engine analyzed [10]. Machine Learningwas discussed in [11].A new advanced indexing was introduced in [12] for efficient IR. But, it did not carry out well through large datasets.A new approach was designed in [13] for biomedical IR based on keyword search. But, it has additional time.NLP-basis of feature weighting scheme[14] introduced. At [15], ensemble scheme designed. To examine query reformulation, graph-basis method employed [16]. Proficient data-basis technique investigated [17].

The remainder section of the paper is arranged in following manner: Section - 2 explains about related research done in this area. The proposed architecture was discussed in Section - 3. The system was compared with existing model in performance of the system in Section - 4. Section - 5 presented the conclusion and future directions.

2. Related Works

A clustering-based method was introduced in [2] with the aim of increasing the average precision and recall in medical information retrieval. However, it failed to design an automatic performance estimation method for medical information retrieval with minimal time. The Exponential Aquila Optimizer (EAO)-based Deep Fuzzy Clustering method, developed in [3], aimed to retrieve relevant documents based on query matching using the Tversky index. However, it faced challenges when applied to larger datasets, impacting its effectiveness in retrieving documents. Rescue provided [18] via similarity matrix mixture. Two-step training plan illustrated [19]. Clinical judgment support method analyzed [20].

A pre-trained BERT model was developed in [21] for healthcare IR. An ontology-based clustering method was designed in [22] for semantic IR. A deep learning-based search was designed in [23] for IR.A two-stage IR method was designed in [24] for biomedical IR. But, it failed to improve retrieval accuracy.

Study focused on biomedical IR [25]. At [26], Contrastivepre-trained transformer employed. Query growth investigated [27] by Optimization. At [28],two-phase IR designed. Rich semantics gathered [29] by multimodal IR. Two-level querymethod performed [30]. Data extraction method proposed [31]. With IR, multi manager basis of approach utilized [32].

3. Methodology

To achieve higher health IR as well as fewer time, GPTCRMCB employed.Proposed contributions given by: For higher precision, GPTCRMCB presented. Employ Gensim tokenizer as well as Gestalt pattern matching with less time. To extract keyword, utilizes targeted projection pursuit. Applicable IR discovered by Tversky regression. Apply Miyaguchi-Preneel cryptographic with higher privacy.

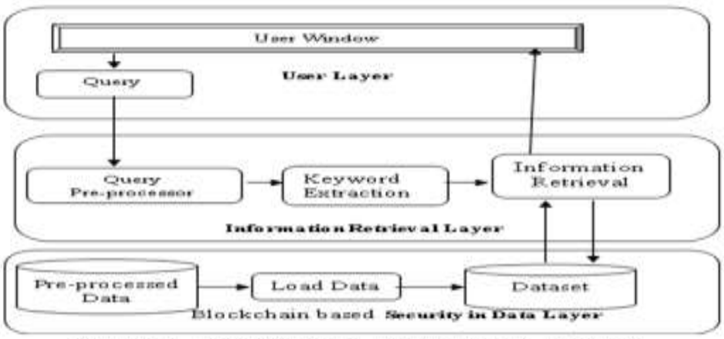

Figure1. Architecture of proposed method

GPTCRMCB is depicted in figure 1 for privacy preserved medical IR.

3.1 preprocessing

Initially, user submits their query into user window and it transmitted to system. After that, preprocessing is executed.It involved tokenization and stop word removal procedures. The Gensim tokenizer is utilizedto segment input query into individual words or tokens, employing punctuation and spaces within the square bracket. Let us consider 'n' number of queries $Q_1, Q_2, Q_3 \dots . Q_n$ and 'm' number of words $w_1, w_2, w_3, \dots . w_m$. Then the tokenization process is expressed as follows,

$$Q \overset{\text{GT}}{\Longrightarrow} [{}'w_1', \ {}'w_2', \ {}'w_3', \dots {}'w_m'] \quad (1)$$

Where, query 'Q' is partitioned into a number of words $w_1, w_2, w_3, \dots . w_m$ using Gensim tokenizer'GT'. Afterward, stop words are detached using gestalt pattern matching technique. It is a statistical technique for analyzing words at predefined file and comparing them with words in current query. Let us consider 'm' number of words or tokens$w_1, w_2, w_3, \dots . w_m$ and predefined stop word list 'w_L'. The matching is performed as follows,

$$PM = 2 * \frac{Mw_s}{|w_m||w_L|} \quad (2)$$

$$PM = \begin{cases} 1, & Stopword \\ 0, & Otherwise \end{cases} \quad (3)$$

Where PMdenotespattern matching score, Mw_s denotes matched words, $|w_m|$ and $|w_L|$ represents number of string in word within query 'w_m' and words in predefined stop word list 'w_L'. The PMprovides output values between zero and one. If $PM$is 1, word is stop words and itdetachedas of query. Algorithm 1 described process of query preprocessing.

// Algorithm 1: Query preprocessing

```
Begin
Collect number of user queries Q_1, Q_2, Q_3 .... Q_n from dataset
  For each query 'Q_i'
     Partition the text into words ['w_1', 'w_2', 'w_3', ... 'w_m'] using tokenizer
   End for
  For each word 'w_i' in query
      For each word 'w_j' in predefined stop word list
         Measure pattern matching score using (2)
  if (PM = 1) then
        Words are matched
    else
        Words are not matched
  End if
    Matched words are identified as stop words and removed
End for
End for
Return (preprocessed words)
End
```

3.2 Targeted projection pursuit technique based keyword extraction

After preprocessing, keywords are extracted throughusing targeted projection pursuit technique.For every word in preprocessed queries, frequent occurrences of words is computed as follow,

$$FO = \frac{NFw}{N_w} \quad (4)$$

Where, FO denotes a frequent occurrences of words, N_w denotes a total number of words in document, NFw denotes a number of words frequently occurred. The projectionfunctionminimizes difference between the frequent occurrences of words and target.

$$Y = \min|T - FO.A| (5)$$

Where, Y denotes an output of target projection, min denotes minimizes difference between target 'T'and frequent occurrences of words 'FO', A indicates a projection matrixto project original keywords. Therefore, keywords are extracted and process shown in algorithm 2.

// Algorithm 2 Targeted projection pursuit based keyword extraction

Begin
Collect Pre-processed user queries
  For each Pre-processed query 'Q_i'
    Compute the frequency occurrence of words using (4)
  End for
Projection function find minimum deviation using (5)
Return (keywords)
End

3.3 Tversky indexive censored regression based information retrieval

Finally, relevant medical information is retrieved using extracted keywordsusing Tversky indexive censored regression. Censored regression is a statistical analysis where the exact value of an output variable is not observed or recorded, and other values are removed due to certain constraints, also known as a censoring point. The censoring point refers to the specific value or threshold that serves as a boundary beyond which the data points are censored. The regression process involves two types of censoring namely left-censoring and right-censoring. In left-censoring, observed value of a Tversky similarity index is known to be less than a certain threshold. Right-censoring occurs when Tversky similarity index output is known to be above a certain threshold.Let us consider number of extractedwords$w_1, w_2, w_3, \dots w_k$.

$$TC = \frac{w_k \cap C_b}{a(w_k \cap C_b) + b(w_k \Delta C_b)} \quad (6)$$

Where, TC indicates a tversky coefficient or relevance score, a and b are parameters of Tversky coefficient (a, b ≥ 0), and coefficient TCprovides output ranging between 0 and 1. $w_k$denotes extracted keywords, C_b denotes a content in database, $w_k \cap C_b$indicates a mutual dependence between keywords and content in database, $w_k \Delta C_b$indicates a variance between keywords and content in its database. Censored regression is a class of models in which relevance score is censored above or below a certain threshold.

$$Z = \begin{cases} TC > \delta\ ; relavant\ information \\ otherwise;\ \ irrelavant\ information \end{cases} \quad (7)$$

Where, Z indicates a censored regression output, δ denotes a threshold, TC denotes a tversky coefficient or relevance score. If $TC$is greater than threshold, relevant information is retrieved. If $TC$is less than threshold, content is considered as irrelevant information and it is censored. Algorithm 3 describes the process of retrieving relevant information.

// Algorithm 3: Tversky indexive censored regression based information retrieval

Begin
Collect $w_1, w_2, w_3, \dots w_k$
  For each keywords 'w_k'
    For each content in database 'C_b'
      Measure relevance score of words using (6)
    End for
  End for
  if ($TC > \delta$) **then**
      Relevant data
        else
      Irrelevant data
  end if
Return (relevant information related to a user query)
End

3.4 Miyaguchi-Preneel cryptographic hash Blockchain for privacy preserved medical IR

After finding relevant information, Miyaguchi-Preneel cryptographic hash Blockchain used to protect privacy of medical data by preventing unauthorized access. Relevant patient information '$D_s = D_1, D_2, \dots D_n$' is considered as input. Input separated to different message agreed sizes blocks.

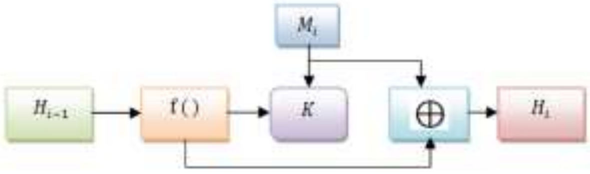

Figure 2 Process of compression function

Compression function illustrated in Figure 2. Input message block 'M_i' encrypted. Prior hash value 'H_{i-1}' set to function f() changed to fit for block cipher 'K'. Output hash generatedwith 'M_i' and XORed with 'H_{i-1}'.

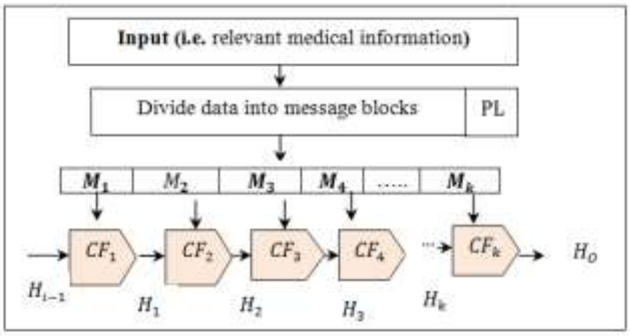

Figure 3 structure of Miyaguchi-Preneel single-block compression

Figure 3 given above depicts structure of Miyaguchi-Preneel compression to generate output hash 'H_O'. Message block indicated as '$M_1, M_2, M_3, M_4, .. M_k$', $CF_1, CF_2, CF_3, CF_4, .. CF_k$ denotes a Miyaguchi–Preneel compression function, last block is padded by '0's (i.e. Pad = 0)and PL denotes Padding length for adding extra bits or bytes to the input data when the input data length is not an exact multiple of block size. By compression, output hashcreated. At first iteration, no H_{i-1} provided as well as employs pre-specified initial value ($H_0 = 0$). Output hash estimated by,

$$H = [K_D(M_i) \oplus H_{i-1} \oplus M_i] \quad (8)$$

Where 'M_i' and 'H_{i-1}'preset to '0', 'K' denotes a block cipher,key to 'K' XORed with 'H_{i-1}' as well as (M_i) denoted as D. Output of last compression function 'H_O' is considered as a final output hash value. In this manner, each applicable medical data is to build hash value.Algorithm 4 describes outlines for privacy aware medical IR from search engine.

// Algorithm 4 Miyaguchi-Preneel cryptographic hash Blockchain

Begin

Collect relevant patient data $D_s \in D_1, D_2, \dots D_n$
For every 't'
Create blockchain
For each 'D_s'
Partitioned to 'k' $D_s \rightarrow M_1, M_2, M_3, \dots M_n$
for both 'M_i'
Produce $H = [K_D(M_i) \oplus H_{i-1} \oplus M_i]$
End for
Obtain 'H_O'
End for
Send hashed data to authorized user
End for

End

4. Experimental Scenario

In this section, GPTCRMCB, Clustering-based Method (CBM)[2], EAO-based Deep Fuzzy Clustering (EAO-DFC) [3],are executedbypython by applying NFCorpus dataset.Performance analysis of proposed system is compared with two existing model CBM and EAO-DFC.It consists of 3,244 NL queries by 169,756 relevance judgments. The dataset contains 9,964 medical documents.

Table 1 **system specification**

Operating system	Windows 10 and above (64-bit)
Language	Python 3.10.11
Hard disk	500 GB
Processor	Intel Pentium processor
RAM	4GB
Dataset	NF Corpus dataset and taken from https://www.cl.uni-heidelberg.de/statnlpgroup/nfcorpus/

Accuracy: User queries percentage in which relevant medical documents correctly retrieved defined as 'Acc'.

$$Acc = \left(\frac{\text{TRP+TRN}}{\text{TRP+TRN+FLP+FLN}}\right) * 100 \quad (9)$$

Where, Accindicates accuracy, true positivesisTRP, true negativeisTRN, false positiveisFLP, false negativeisFLN. Accuracy is measured in percentage.

Precision (PR):it used atdata retrieval that measures proportion of relevant documents related to query.

$$PR = \frac{\text{TRP}}{\text{TRP+FLP}} \quad (10)$$

Recall (RL): it measures ratio of pertinent medical records correctly retrieved.

$$RL = \frac{\text{TRP}}{\text{TRP+FLN}} \quad (11)$$

F1-score: It mergesPRas well as RL rate.

$$F1 - score = 2 * \frac{PR*RL}{PR+RL} \quad (12)$$

Information retrieval time: time neededfor relevant or irrelevant medical documents based on user queries referred as 'IRT'.

$$IRT = \sum_{i=1}^{n} Q_i * Time\ [RD] \qquad (13)$$

Where, IRTindicates Information retrieval time,$Time\ [RD]$ indicates a time for retrieving relevant document 'RD' based on user queries 'Q_i'. The time is calculated in milliseconds (ms).

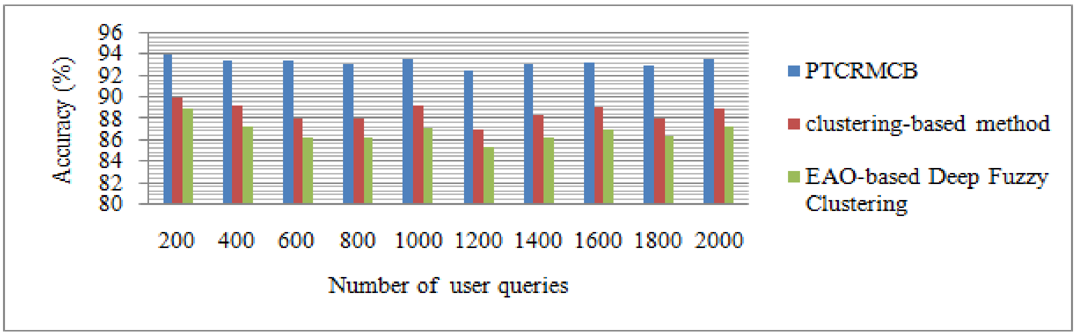

Figure 4 Accuracy

Figure 4depicts experimental results of accuracy versus queries, ranging from 200 to 2000. The results of GPTCRMCBmethod is increased by5% and 7% compared to [CBM] [EAO-DFC]respectively.

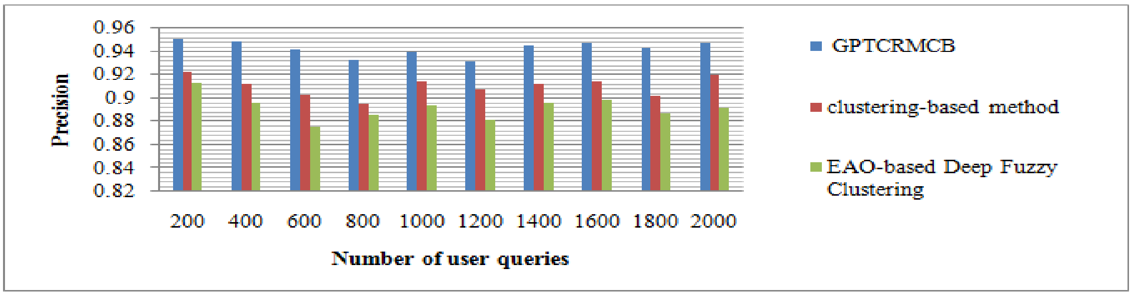

Figure 5 Precision

Figure 5 depicts a performance analysis of precision versus number of user queries. The GPTCRMCBimproved the precision by 4% compared to [CBM] and 6% compared to [EAO-DFC].

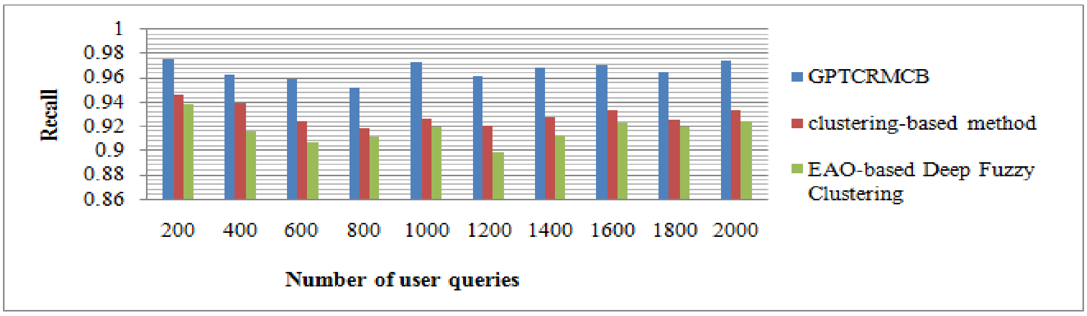

Figure 6 Recall

Figure 6 displays performance results of recall. The results indicate the GPTCRMCB method is found to be higher by 4% and5%than the[CBM] and [EAO-DFC] respectively.

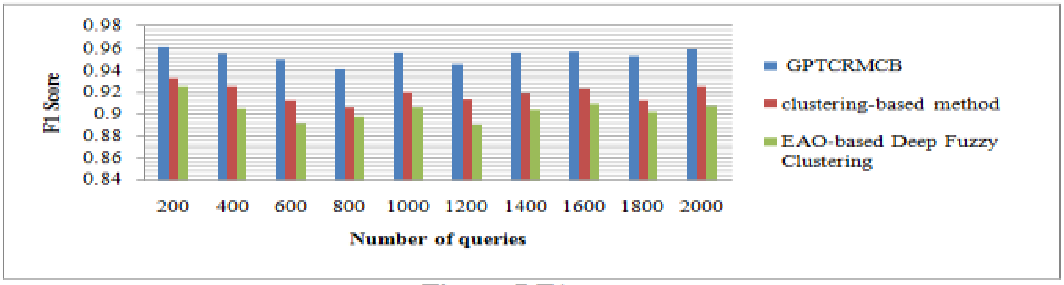

Figure 7 F1-score

The F1-score for retrieving relevant medical information are depicted in Figure 7. Outcomes of GPTCRMCB are improved by 4%, 6%compared to [CBM] and[EAO-DFC] respectively.

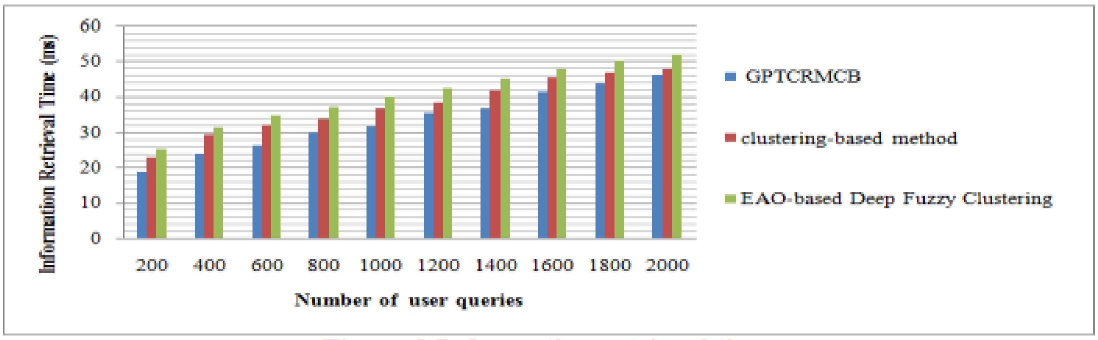

Figure 8 Information retrieval time

Figure 8depicts performance outcomes of IR time using three different methods.GTPPTICR of IRT minimized by 12%, 18%compared to [CBM], [EAO-DFC]respectively.

Proposed GPTCRMCB and[CBM], [EAO-DFC] are implemented using python by applying 500 GB dataset for evaluating clinical data retrieval schemes as exposed in figure 9 and 10.

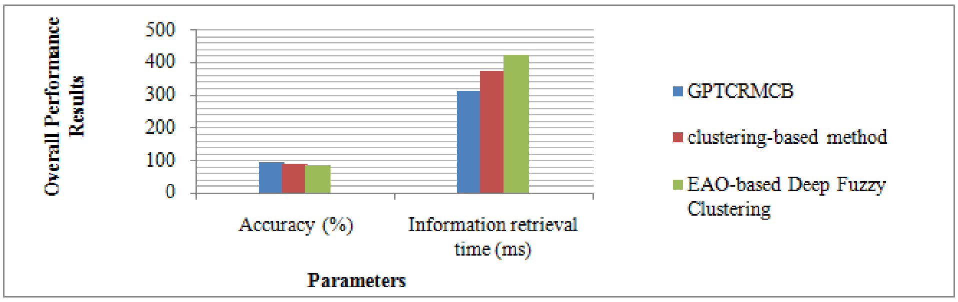

Figure 9 graphical representations of accuracy and Information retrieval time

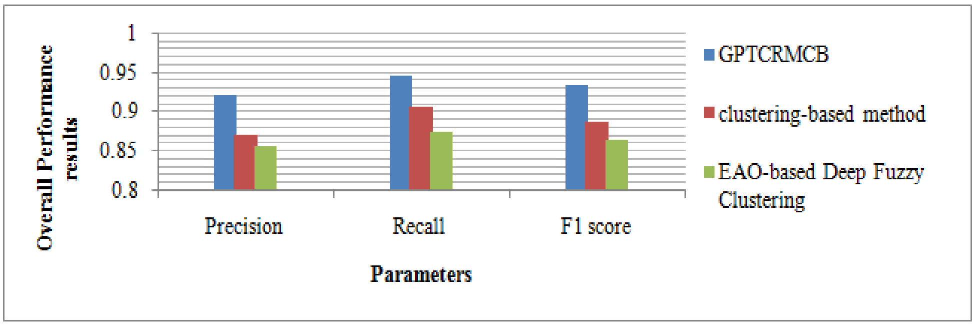

Figure 10 graphical representations of precision, recall, F1 Score

Figure 9 and 10 illustrates performance results of three parameters using 500GB dataset. The performance of all parameters is improved using GPTCRMCB than existing [CBM] [EAO-DFC].

5. Conclusion

In this paper, a novel GPTCRMCB has designed for accurate IR. The GPTCRMCB method performs query preprocessing and keyword extraction to minimize time. The Censored Regression is employed to retrieve relevant patient information. Finally, cryptographic hash Blockchain is employed for secure IR. The paper conducts an experimental evaluation using NFCorpus dataset and 500GB dataset for Medical IR. The observed results show the GPTCRMCB method increases theperformance compared to existing approaches.However, GPTCRMCB method faced demands when dealing throughmultifaceted queries, especially in contextual perception based keyword matching. In the future, advanced deep learning models will be used for query preprocessing to handle complex and context-aware patient IR.

References

1. Keejun Han and Jungeun Kim, “Proximity-Aware Clinical Passage Retrieval Framework by Exploiting Knowledge Structure”, IEEE Access, Volume 11, 2023, Pages 37681 - 37693. DOI: 10.1109/ACCESS.2023.3266004
2. Qiuyu Xu, Yidong Huang, Shengli Wu, Chris Nugent, “Clustering-based fusion for medical information retrieval”, Journal of Biomedical Informatics, Elsevier, Volume 135, November 2022, Pages 1-12. https://doi.org/10.1016/j.jbi.2022.104213
3. Gunjan Chandwani, Dr. Anil Ahlawat, Dr. Gaurav Dubey, “Aspect term extraction and optimized deep fuzzy clustering-based inverted indexing for document retrieval”, Intelligent Decision Technologies, Volume 16, Issue 3, 2022, Pages 543-555. https://doi.org/10.21203/rs.3.rs-1660924/v1
4. Hanan Naser Alsuqaih, Walaa Hamdan, Haythem Elmessiry , Hussein Abulkasim, “An efficient privacy-preserving control mechanism based on blockchain for E-health applications”, Alexandria Engineering Journal, Elsevier, Volume 73, 2023, Pages 159-172. https://doi.org/10.1016/j.aej.2023.04.037

5. L. Arokia Jesu Prabhu , Sudhakar Sengan , G.K. Kamalam , J. Vellingiri , Jagadeesh Gopal , Priya Velayutham , V. Subramaniyaswamy, "Medical Information Retrieval Systems for e-Health Care Records using Fuzzy Based Machine Learning Model", Microprocessors and Microsystems, Elsevier, 2020, Pages 1-28. https://doi.org/10.1016/j.micpro.2020.103344
6. Xia Ning, Ziwei Fan , Evan Burgun, Zhiyun Ren, Titus Schleyer, "Improving information retrieval from electronic health records using dynamic and multi-collaborative filtering", PLoS ONE, Volume 16, Issue 8, 2021, Pages 1-24. https://doi.org/10.1371/journal.pone.0255467
7. Mr.Kshirsagar Sopan B ,Mr.Gawade Nilesh M., "Data Retrieval from Electronic Health Records", International Journal of Creative Research Thoughts, Volume 9, Issue 1, 2021, Pages 96-98. http://www.ijcrt.org/viewfull.php?&p_id=IJCRTF020021
8. Yun Dai and Hao Liu, "Application of the R-Tree Clustering Model in Medical Information Retrieval", Mobile Information Systems, Hindawi, Volume 2022, August 2022, Pages 1-9. https://doi.org/10.1155/2022/2887395
9. Poly Sil Sen & Nandini Mukherjee, "An ontology-based approach to designing a NoSQL database for semi-structured and unstructured health data", Cluster Computing, Springer, Volume 27, 2024, Pages 959–976. https://doi.org/10.1007/s10586-023-03995-y
10. David A. Hanauer, Jill S. Barnholtz-Sloan, Mark F. Beno, Guilherme Del Fiol, Eric B. Durbin, Oksana Gologorskaya, Daniel Harris, Brett Harnett, Kensaku Kawamoto, Benjamin May, Emily Pfaff, Janie Weiss, and Kai Zheng, "Electronic Medical Record Search Engine (EMERSE): An Information Retrieval Tool for Supporting Cancer Research", JCO Clinical Cancer Informatics, Volume 4, 2020, Pages 454- 463. https://doi.org/10.1200/CCI.19.00134
11. Maria Frasca & Genoveffa Tortora, "Visualizing correlations among Parkinson biomedical data through information retrieval and machine learning techniques", Volume 81, 2022, Pages 14685–14703. https://doi.org/10.1007/s11042-021-10506-x
12. Doaa N. Mhawi, Haider W. Oleiwi, Nagham H. Saeed and Heba L. Al-Taie, "An Efficient Information Retrieval System Using Evolutionary Algorithms", Network, Volume 2, Issue 4, 2022, Pages 583-605. https://doi.org/10.3390/network2040034
13. Shariq Bashir, Akmal Saeed Khattak, Mohammed Ali Alshar, "Automatically transforming full length biomedical articles into search queries for retrieving related articles", Egyptian Informatics Journal, Springer, Volume 22, Issue 1, March 2021, Pages 75-84. https://doi.org/10.1016/j.eij.2020.04.004
14. Jainisha Sankhavara, "Feature Weighting in Finding Feedback Documents for Query Expansion in Biomedical Document Retrieval", SN Computer Science, Springer, Volume 1, 2020, Pages 1-7. https://doi.org/10.1007/s42979-020-0069-x
15. Zhiqiang Liu, Jingkun Feng, Zhihao Yang, Lei Wang, "Document Retrieval for Precision Medicine Using a Deep Learning Ensemble Method", JMIR Medical Informatics, Volume 9, Issue 6 , 2021, Pages 1-12. DOI: 10.2196/28272
16. Jainisha Sankhavara, Rishi Dave, Bhargav Dave, Prasenjit Majumder, "Query specific graph-based query reformulation using UMLS for clinical information access", Journal of Biomedical Informatics, Elsevier, Volume 108, 2020, Pages 1-8. https://doi.org/10.1016/j.jbi.2020.103493
17. Stefano Marchesin, Giorgio Maria Di Nunzio and Maristella Agosti, "Simple but Effective Knowledge-Based Query Reformulations for Precision Medicine Retrieval", Information, Volume 12, Issue 10, 2021, Pages 1-27. https://doi.org/10.3390/info12100402
18. Andrés Rosso-Mateus, Manuel Montes-y-Gómez, Paolo Rosso and Fabio A. González, "Deep fusion of multiple term-similarity measures for biomedical passage retrieval", Journal of Intelligent & Fuzzy Systems, Volume 39, Issue 2, 2020, Pages 2239-2248. **DOI:** 10.3233/JIFS-179887
19. Wojciech Kusa, Óscar E. Mendoza, Petr Knoth, Gabriella Pasi, Allan Hanbury, "Effective matching of patients to clinical trials using entity extraction and neural re-ranking", Journal of Biomedical Informatics, Elsevier, Volume 144, 2023, Pages 1-12. https://doi.org/10.1016/j.jbi.2023.104444
20. Zhengyun Zhao, Qiao Jin, Fangyuan Chen, Tuorui Peng & Sheng Yu, "A large-scale dataset of patient summaries for retrieval-based clinical decision support systems", Scientific Data, Volume 10, 2023, Pages 1-14.https://doi.org/10.1038/s41597-023-02814-8
21. Sameh Frihat, Catharina Lena Beckmann, Eva Maria Hartmann and Norbert Fuhr, "Document Difficulty Aspects for Medical Practitioners: Enhancing Information Retrieval in Personalized Search Engines", Applied Sciences, Volume 13, Issue 19, 2023, Pages 1-15. https://doi.org/10.3390/app131910612
22. Maryam Shakeri, Abolghasem Sadeghi-Niaraki, Soo-Mi Choi and Tamer AbuHmed, "AR Search Engine: Semantic Information Retrieval for Augmented Reality Domain", Sustainability, Volume14, Issue 23, 2022, Pages 1-16. https://doi.org/10.3390/su142315681
23. Andre Esteva, Anuprit Kale, Romain Paulus, Kazuma Hashimoto, Wenpeng Yin, Dragomir Radev & Richard Socher, "COVID-19 information retrieval with deep-learning based semantic search, question answering,

and abstractive summarization", npj Digital Medicine, volume 4, 2021, Pages 1-9. https://doi.org/10.1038/s41746-021-00437-0
24. Zicheng Zhang, Xinyue Lin and Shanshan Wu, "A hybrid algorithm for clinical decision support in precision medicine based on machine learning", BMC Bioinformatics, Volume 24, Issue 3, 2023, Pages 1-18. DOI: 10.1186/s12859-022-05116-9
25. Pir Noman Ahmad, Yuanchao Liu, Khalid Khan, Tao Jiang and Umama Burhan , "BIR: Biomedical Information Retrieval System for Cancer Treatment in Electronic Health Record Using Transformers", Sensors, Volume23, Issue 23, 2023, Pages 1-23. https://doi.org/10.3390/s23239355
26. Qiao Jin, Won Kim, Qingyu Chen, Donald C Comeau, Lana Yeganova, W John Wilbur, Zhiyong Lu, "MedCPT: Contrastive Pre-trained Transformers with large-scale PubMed search logs for zero-shot biomedical information retrieval", Bioinformatics, Volume 39, Issue 11, 2023, Pages 1-6. https://doi.org/10.1093/bioinformatics/btad651
27. Ram Kumar, S. C. Sharma, "Hybrid optimization and ontology-based semantic model for efcient text-based information retrieval", The Journal of Supercomputing, Springer, Volume 79, 2023, Pages 2251–2280. https://doi.org/10.1007/s11227-022-04708-9
28. Hao Cong, Wei-Neng Chen & Wei-Jie Yu, "A two-stage information retrieval system based on interactive multimodal genetic algorithm for query weight optimization", Complex & Intelligent Systems, Springer, Volume 7, 2021, Pages 2765–2781. https://doi.org/10.1007/s40747-021-00450-6
29. Chi Thang Duong, Thanh Tam Nguyen, Hongzhi Yin, Matthias Weidlich, Thai Son Mai , Karl Aberer, Quoc Viet Hung Nguyen, "Efficient and Effective Multi-Modal Queries Through Heterogeneous Network Embedding", IEEE Transactions on Knowledge and Data Engineering , Volume 34, Issue 11, 2022, Pages 5307 – 5320. **DOI:** 10.1109/TKDE.2021.3052871
30. Namrata Rastogi, Parul Verma, Pankaj Kumar, "Query Expansion based on Word Embeddings and Ontologies for Efficient Information Retrieval", International Journal of Advanced Computer Science and Applications, Volume 12, issue 11, 2021, Pages 367-373. DOI:10.14569/IJACSA.2021.0121142
31. Mohamed Yassine Landolsi, Lobna Hlaoua& Lotfi Ben Romdhane, "Information extraction from electronic medical documents: state of the art and future research directions", Knowledge and Information Systems, Springer, Volume 65, 2023, Pages 463–516. https://doi.org/10.1007/s10115-022-01779-1
32. Abdel Naser Pouamoun, Ilker Kocabas, "Multi-agent-based hybrid peer-to-peer system for distributed information retrieval", Journal of Information Science, Volume 49, Issue 3, 2021, Pages 1-15.DOI:10.1177/01655515211010392.
33. Nidhi Mishra. (2025). Integration of Hybrid Energy Storage Systems with Smart Grids: A Control-Oriented Modeling and Experimental Validation Study. *Transactions on Energy Storage Systems and Innovation,* 34-41.
34. A. Surendar. (2025). Context-Adaptive Intelligent Window-Integrated Metasurfaces for Resilient Millimetre-Wave Signal Propagation in Smart Environments. *National Journal of Ubiquitous Computing and Intelligent Environments*, *2*(3), 28–35.
35. Robbi Rahim. (2026). Bio-Computational Simulation of Microbial Invasion Dynamics: Implications for Biosecurity and Rural Environmental Protection. *National Journal of Smart Agriculture and Rural Innovation*, *4*(2), 9-15.
36. Nikola Jovanović, Maja Petrović, & Marko Ilić. (2025). Building Excellence in Education through Evidence-Based Practice. *National Journal of Quality, Innovation, and Business Excellence*, *2*(2), 12–23.