RAG-Driven Drug Recommendations: A Novel Approach to Personalized Medicine

Kritika Tripathi[1], Devanshi Malik[2] and Sheenam Naaz[3]
[1,2,3]*Department of Computer Science Engineering, Sharda University, Greater Noida, India*

Abstract*— Drug recommendation systems have the potential to significantly improve personalized medicine by tailoring therapeutic options to individual patient profiles and medical histories. However, these systems often face challenges in accurately synthesizing large, diverse sources of biomedical information. To improve personalized medicine, this paper introduces a drug recommendation framework powered by using Retrieval-Augmented Generation (RAG) technique. By combining a retrieval mechanism and a generative model, the system effectively combines patient data with a rich biological literature to provide specific medication recommendations. Based on experimental results, the RAG-based approach not only increases the performance of recommendations but also enhances interpretability by informing doctors about relevant treatment choices and why each recommendation came from. This method appears the feasible way for doing precision medicine and informed clinical performance in numerous healthcare situation.

***Keywords*—drug recommendation, machine learning, NLP, Retrieval-Augmented Generation**

I. INTRODUCTION

The application of artificial intelligence (AI) and machine learning (ML) techniques is changing clinical decision-making in the field of modern healthcare, especially in the area of drug recommendation systems. By prescribing individualized treatments based on patient-specific data, including demographics, medical history, genetic variables, and drug interactions, these systems significantly contribute to the guidance of healthcare professionals. There is a greater need than ever for complex algorithms that can traverse and make sense of the ever-increasing amount of healthcare data that is available. Precise medication recommendations contribute to better treatment outcomes and lessen adverse drug reactions, which are a major global source of hospitalizations and medical expenses.

The healthcare industry has seen significant advancements in recent years, particularly in the development of intelligent systems for personalized medicine. One of the key areas of focus has been drug recommendation systems, which are designed to provide personalized therapeutic options based on individual patient profiles. These systems can be highly valuable, as they have the potential to optimize treatment strategies, minimize adverse drug reactions, and improve overall patient outcomes. However, despite significant progress, traditional drug recommendation methods still face numerous challenges, particularly in their ability to synthesize and process the ever-growing amount of diverse biomedical information. Furthermore, many existing models lack the ability to generate dynamic, context-sensitive recommendations that can adjust based on new or changing patient data.

Drug recommendation systems generally rely on two primary approaches: rule-based systems and machine learning-based systems. Rule-based systems use predefined criteria, such as drug interactions, patient characteristics, and known treatment guidelines, to make recommendations. While these systems are relatively simple to implement, they often lack the flexibility and adaptability needed to handle the complexities and nuances of patient data. On the other hand, machine learning-based systems, including collaborative filtering and content-based models, attempt to learn patterns in data to recommend drugs based on patient profiles or similar patient outcomes. While these systems have shown promise, they often struggle with generalization and can be limited by the quality and breadth of the training data.

With the increasing availability of large-scale biomedical literature, electronic health records (EHR), and clinical guidelines, there is a growing need for systems that can effectively integrate these diverse information sources. Moreover, the ability to explain and justify recommendations in a manner that clinicians can understand and trust is critical in healthcare applications. Recent advancements in natural language processing (NLP) and deep learning have made it possible to address these challenges more effectively. Specifically, the advent of Retrieval-Augmented Generation (RAG) models has shown great promise in combining retrieval-based and generative capabilities, providing a way to bridge the gap between traditional recommendation systems and more complex, dynamic approaches.

RAG[29] is a hybrid model that integrates a retrieval mechanism with a generative model, enabling the system to first retrieve relevant information from a large corpus of data and then generate a tailored response based on that retrieved information. This hybrid approach allows the model to leverage the vast wealth of existing biomedical knowledge, including scientific publications, clinical trial results, and patient-specific data, while also generating novel recommendations that consider the unique context of each individual. By combining the strengths of both retrieval and generation, RAG models offer a more flexible and adaptive approach to drug recommendation, allowing for real-time, context-aware suggestions that are both accurate and interpretable.

In this paper, we present a drug recommendation system based on Retrieval-Augmented Generation (RAG). The proposed system combines biomedical literature, clinical knowledge, and patient data to retrieve relevant drug information, and then generates personalized recommendations that are tailored to the individual patient's needs. We demonstrate the effectiveness of this approach by evaluating its performance on benchmark datasets and comparing it to traditional recommendation systems. Our results show that the RAG-based model significantly improves recommendation accuracy and provides higher interpretability compared to existing systems, making it a promising tool for clinical decision support and precision medicine[30].

This paper proposes a RAG-based drug recommendation framework that integrates both retrieval and generation for more accurate and context-sensitive recommendations. This approach has the potential to revolutionize clinical decision-making by offering a more reliable and transparent means of recommending drugs that are not only suitable for the patient but also supported by a wealth of evidence from the medical literature.

LITERATURE SURVEY

K. P. Swain et al. [1] proposes a system that harnesses machine learning and patient feedback to enhance drug recommendations, aiming to minimize medical errors by providing personalized suggestions based on historical data and patient experiences. They have used a blend of machine learning strategies and Natural Language Processing (NLP) techniques to sift through extensive pharmaceutical data and aims to reveal latest patterns and trends which are crucial for assessing drug performance.

R. Ramachandran et al. [2] proposed a new AI-based learning method which is the Artificial Intelligence based Neural Drug Suggestion Model (AINDSM). It assesses the effectiveness of the scheme using a cross-validation with the traditional Support Vector Machine (SVM). They provided a medication recommendation tools to make infectious illness treatment fair and safe.

G K Prashanth et al. [3] proposed a medicine recommendation system to lessen the backlog of specialists by using vectorization techniques including Bow, TF-IDF, Word2Vec, and Manual Feature Analysis. This proposed system gives an accuracy of 97% with a comprehensive dataset comprising patient demographics, genetic information, medical history, and existing medical records. They used Naïve Bays machine learning model to train in this data by extracting relevant features using feature engineering techniques.

K. C Sreedhar et al. [4] proposed a drug suggestion system using emotion analysis of patient reviews and machine learning. They have used various vectorization techniques to suggest the best drug and used classification methods such as LinearSVC, and accuracy, F1-score, precision, AUC score to rate their emotions. The results achieved 93% accuracy using LinearSVC along with TF-IDF vectorization.

Samuthira Pandi et al. [5] proposed a novel Patient Drug Recommendation System (PDRS) using Cascaded Multi-Layer Perceptrons (MLP). They used extensive Drug Review Dataset to recommend precise and personalized drug recommendations by sequentially processing and analyzing diverse aspects of drug reviews, capturing nuanced information using cascaded MLP. The system provides a remarkable accuracy of 97%.

Xiang Li et al. [6] introduced a model to enhance patient's personalization by focusing on patient condition which is DisMed. In this system, they establish a correlation between conditions and medications by examining patient's condition in depth and recognizing and adapting to the evolving nuances of the conditions to map them to corresponding medications. They further evaluate on real-world datasets and demonstrated that they surpass leading models and improves the customization of patient profiles.

B. Sivaiah et al. [7] proposed a system to analyse sentiments using advanced vectorization methods to identify subtle sentiment nuances and classification algorithms such as Naïve Bayes, SVC, Random Forest and LinearSVC which is crucial for tailoring personalized drug recommendations. The LinearSVC classifier coupled with TF-IDF vectorization provided superior sentiment prediction performance and enhance healthcare practices.

Kornpong Ariyachaipong et al. [8] proposed drug recommendation system to ensure high-quality recommendations using drug details and a combination of cosine similarity and Bee algorithm. A baseline for drug similarity is established using cosine similarity and further the selection process is optimized using Bee algorithm by considering additional factors.

Kadime Göğebakan et al. [9] proposed a drug prescription recommendation system using DIAKID ontology and extensive semantic rules for Type-2 Diabetes Mellitus (T2DM) and CKD. They used comprehensive Semantic Web Rule Language (SWRL) rules to extract the correct drug dose, K-raising drugs, and drug-drug interaction warnings based on the Glomerular Filtration Rate (GFR) value of T2DM and CKD patients. The proposed work achieves competitive results and guide to prepare prescriptions by giving warnings about drug-drug interactions and doses.

Sinchana N M et al. [10] proposed a drug recommendation system that utilizes advanced algorithms and techniques and model-based filtering techniques such as Singular Value Decomposition (SVD) and Non-Negative Matrix Factorization (NMF) to recommend the right drug. They used large diabetes patient datasets to enhance the accuracy and effectiveness of drug recommendations. They evaluated the system on three different standard evaluation metrices which are bar graph which demonstrates the superiority of SVD over NMF.

Sawrawit Chairat et al. [11] proposed an approach to recommend drug which involves word embedding to transform unstructured outpatient records into dense numerical vectors and further used hierarchical recurrent neural network to make hierarchical predictions at four different levels of Anatomical Therapeutic Chemical (ATC) medication codes. They achieved F1 scores of 0.7128, 0.6475, 0.3834 and 0.3156 at the 1^{st}, 2^{nd}, 3^{rd}, and 4^{th} levels.

Qasem Kharma et al. [12] proposed a drug recommendation model that effectively offer more structured and granular representation of drugs. This model uses taxonomy and multi-criteria collaborative filtering, it outperforms baseline approach, improves prediction accuracy.

Junping Liu et al. [13] proposed FDIRNet model to recommend the drug by employing forward imputation to deal with missing and abnormal data. This system utilizes recurrent neural network and residual neural network to capture the changing patterns of a patient's historical information. A novel loss function is proposed, which is combined with a Drug-Drug Interaction (DDI) loss strategy to mitigate DDI, thereby improving the overall robustness of the model. This system provides a result of 3.1% in Jaccard, 1.6% in F1-score and 3.1% in PRAUC.

Smt. Gousiya Begum et al. [14] proposed a system for drug recommendation which uses Natural Language Processing (NLP) and Recurrent Neural Network (RNN). In this, NLP works best for gathering useful information from patients data and RNN works well in analysing textual data. The system is further evaluated on various metrices such as precision, recall, accuracy, f1 score , roc curve.

This study [15] proposed a selective coverage mechanism and adaptive memory neural network. The model uses neural memory network to store the temporal pattern encoding results corresponding to the patient's health status and uses the coverage mechanism to perform data filtering and attention weight adjustment in the iterative reading process. At the same time, the model is based on the patient's situation, adaptively determining the number of readings of the neural memory network and effectively constructing a representation vector of the patient's health status.

C. Silpa et al. [16] proposed a drug recommendation system using decision tree which gives the patient reliable information about the medication, the dosage, and any possible adverse effects. In this, patient's symptoms, blood pressure, diabetes, temperature, and other parameters comes into force to decide about the medication. They suggest that a drug recommendation system is very helpful in times of medical emergency for giving patients recommendations for safe medications.

This study [17] proposed drug recommendation system using collaborative filtering and K-means clustering algorithm which will be used for drug recommendation system for patients in the treatment stage of Covid-19. This recommender system gives acceptable performance based on the evaluation criteria of precision, recall and F1-score.

This study [18] proposed a comprehensive pharmaceutical recommendation system called RECOMMED based on the patients and drugs features extracted from Drugs.com and Druglib.com. This approach used artificial intelligence (AI) model for the implementation. Sentiment analysis is done using natural language processing approaches along with two models are constructed using patients conditions and drugs features based on matrix factorization. As a result, a deep learning model is constructed for recommendation using 2304 patients as a training set and 660 patients as validation set which is the combined with knowledge-based system with the rules obtained from constraints.

Theresa Omodunbi et al. [19] proposed drug recommendation system that will consider patients peculiarities and then recommend. They review the state-of-the-art algorithms applied to drug recommendation system (DRS). They showed that most recent DRSs uses machine learning based algorithms and suggest that most DRS don't use pre-existing medical conditions and further proposed a recommendation system which will also incorporate a feedback mechanism strengthen the knowledge base of the system.

Hadi Al Mubasher et al. [20] proposed a hybrid recommender system using Artificial Neural Networks and Natural Language Processing. The proposed model is evaluated and compared with similar model, the proposed model has advantage over the other similar models.

Gunakar Koganti et al. [21] proposed a patient review-based recommendation system that uses sentiment analysis to assess a drug's efficacy. High accuracy sentiment classification is achieved by combining Naive Bayes with NLP techniques such as TF-IDF.

The study [22] proposed uses Naive Bayes and natural language processing and develops a system to evaluate drug reviews and suggest medications. It makes trustworthy suggestions by classifying reviews into positive or negative feelings using information like N-grams.

This study [23] incorporates medication recommendations and utilises symptoms as input for a disease prediction algorithm. Using natural language processing (NLP), the Naive Bayes classifier classifies symptoms to provide tailored medication recommendations.

The quality and dependability of medication recommendations based on user sentiment are the main topics of this study [24], which uses Naive Bayes and natural language processing to categorise user evaluations. It improves sentiment classification accuracy by using TF-IDF and N-grams.

This study [25] uses LightGBM and Naive Bayes to categorise reviews and suggest medications. Accurately recording user sentiment is made easier by the incorporation of NLP techniques like Word2Vec.

II. DATASET USED

This research uses medical_meadow_medqa dataset, which is a question-answering dataset with around 10200 rows. This dataset contains three columns, the first one is the input, which contains a question and provides four options to choose an answer from, the second column contains the instruction contains the instruction and the third column contains the actual output of the question.

As LLaMA is a pre-trained model, and it is already trained on a large amount of training dataset, it needs to be fine-tuned with some medical dataset that will specialise it for medical question-answering. For this purpose, this research uses the open-source dataset, medical_meadow_medqa.

III. METHODOLOGY USED

In natural language processing (NLP), retrieval-augmented generation (RAG)[26] is a hybrid technique that enhances the precision and quality of language production tasks by fusing generative and retrieval-based models. It was first presented as a solution to some of the drawbacks of conventional generative models, including their need on substantial training data and their incapacity to access external, real-time knowledge.

The aim of RAG(Figure *1*) is to enhance the capabilities of LLM with external knowledge before generating the output. There are two phases of RAG[27]. The first phase is retrieval, in this phase a query is input by the user to the retriebal model that searches for the most relevant documents through the database for the query. It ranks the documents according to their relevance and then outputs the k most relevant documents that could be used for making decisions, this is the external information, that might not be captured by the model's parameters alone.

After this, all the relevant documents are passed to the generative model, which takes in the documents and the query to generate more informed and contextually appropriate outputs. In conclusion, the external information provided to the model by the retrieval model and the model's internal information are combined together to provide an accurate result. This allows the model to generate outputs that are not just limited to its own knowledge, but dynamically incorporated to it as well.

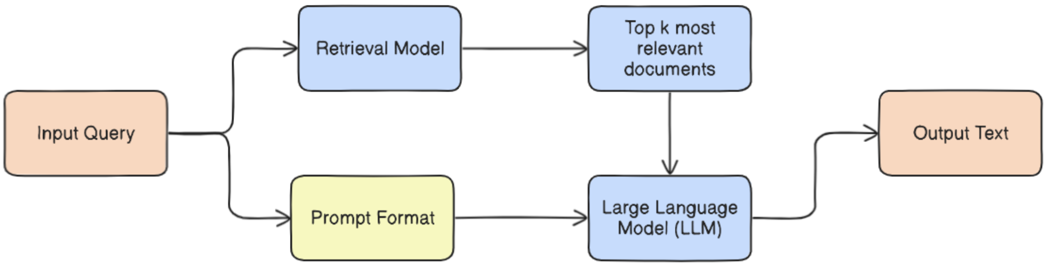

Figure 1. Working of the Retrieval-Augmented Generation Model

A. *Data Preprocessing*

The dataset contains three columns, namely, text input, instruction and output. Because it helps divide lengthy texts into digestible chunks, chunking documents for RAG ensures more accurate retrieval of pertinent information and more efficient processing. The text is normalized and vectorization is performed using text embedding framework. It enables dense embeddings to be extracted from the input text and used in vector similarity search. This dataset is stored in the form of key-value pairs, like a dictionary. The database encodes the user's query into the same embedding space as the stored chunks using the same embeddings method. Figure *2* demonstrates the process of data preparation.

B. *Retrieval*

The system retrieves pertinent documents or information during the retrieval phase in response to user or system queries, such as "Which drug is effective for anxiety?" This is accomplished by:

- Query Encoding: The same methodology used for indexing is applied to translate the user's query into vector space.
- Nearest Neighbor Search: Using closest neighbor search strategies like cosine similarity or dot-product, the encoded query is compared to the indexed corpus to get the top k pertinent items.

Numerous retrieval techniques can be investigated, including:

- Dense retrieval is the process of locating semantically related documents using pre-trained transformer models, such as Bi-Encoder or Multi-Encoder.
- Sparse Retrieval: If the dataset is big and text-heavy, using conventional methods for keyword-based search, such as TF-IDF or BM25.

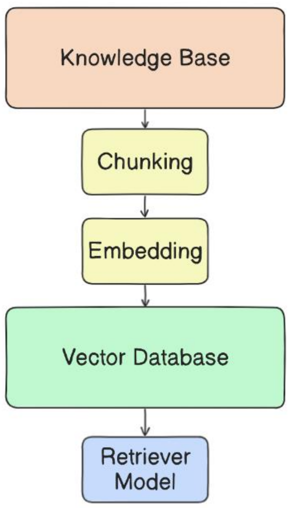

Figure 2. Dataset Preparation

C. Generation

After pertinent materials have been collected, the Retrieval-Augmented Generation (RAG) approach's Generation Phase is essential to producing meaningful, contextually relevant responses[28]. This step combines the external knowledge with the model's prior knowledge to produce a response after the system has acquired the most important information (i.e., documents or data related to the query).

This stage creates a rich input representation for the generative model by combining the user's query with the documents that were retrieved. The input includes:

1) Query Encoding:

Usually, a pre-trained language model like BERT, T5, or a biomedical-specific variation like BioBERT is used to encode the user's query into a dense vector space. This is important because, in order to guarantee that pertinent and context-sensitive material is produced, the generative model must first comprehend the semantics of the query.

2) Document Encoding:

Similar embedding techniques are also used to encode the documents that were retrieved in the previous phase. The most pertinent information is extracted from these documents through processing (e.g., specific pharmacological information or disease treatment alternatives).

3) Concatenation of Query and Context:

A combined input is created by concatenating the query and the documents that were retrieved. The generation phase is set against the backdrop of this combined representation. For instance, the retrieved materials about diabetes treatment and medication efficacy will be coupled with the query, "Which drug is effective for diabetes?"

D. Final Output and Presentation

The final product is often a synthesized drug recommendation or medical advise that is generated and shown in a user-friendly style. This could include a list of medications along with comprehensive details about their interactions, side effects, effectiveness, and other pertinent medical information. The results can be displayed using:

- Text-based recommendations: A detailed output that provides an explanation of the rationale for a certain medication's suggestion based on the documents that were collected.
- Interactive Interfaces: By integrating with Electronic Health Records (EHRs), interactive Q&A systems or suggestion dashboards can provide patients or physicians with individualized treatment alternatives.

IV. Results

The model recommends a drug on the basis of the query provided by the user. In evaluating the drug recommendation system using Retrieval-Augmented Generation (RAG), multiple performance metrics were assessed to determine the model's effectiveness in generating accurate, context-aware, and clinically relevant recommendations. The results are presented across three key aspects: retrieval accuracy, generation quality, and clinical relevance.
The outputs of one input query can be seen in the Figure *3*.

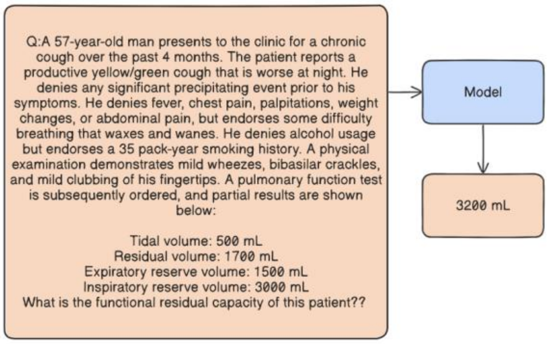

Figure 3. Output from the RAG model

V. CONCLUSION

Encoding the query and retrieved documents, optimizing generative models on domain-specific data, utilizing sophisticated decoding techniques, and making sure the output is tailored and medically valid are all part of the intricate generation phase of a RAG-based drug recommendation system. Clinical decision support and healthcare applications can benefit greatly from RAG's ability to produce highly accurate, context-aware, and tailored medicine recommendations by skillfully fusing retrieval and generative capabilities. In future, this system can be enhanced and built into an interactive chatbot that will recommend medicines and also answer the user queries regarding the medicines and treatments.

REFERENCES

[1] K., P., Swain., Sonali, Mohapatra., Vinayakumar, Ravi., Soumya, Ranjan, Nayak., Tahani, Jaser, Alahmadi., Prabhishek, Singh., Manoj, Diwakar. (2024). Leveraging Machine Learning and Patient Reviews for Developing a Drug Recommendation System to Reduce Medical Errors. The Open Bioinformatics Journal, 17(1) doi: 10.2174/0118750362291402240621044046

[2] R., Ramachandran., V., Chinnammal., K, Malathy., K., Sailaja., S., Sathish, Kumar. (2024). An Improved Drug Recommendation System Using Artificial Intelligence Assisted Learning Methodology. doi: 10.1109/accai61061.2024.10602473

[3] G, K, Prashanth., H, S, Yashaswini. (2024). Drug Recommendation System Using Machine Learning. doi: 10.1109/icsses62373.2024.10561403

[4] K., C, Sreedhar., Bhoomika, Dubba., T., Sowmya, sri., M.Sowbhagya, Pradhaini. (2024). Drug Recommendation System Based on Sentiment Analysis of Drug Reviews using Machine Learning. International Journal For Multidisciplinary Research, doi: 10.36948/ijfmr.2024.v06i03.18767

[5] Samuthira, Pandi., V., Kavitha., S, Arumai, Shinev., Roopa, Ravindranath., A.S., Valarmathy., E, Glory. (2024). Empowering Patient Care: A Novel Patient Drug Recommendation System Using Artificial Intelligence with Modified Learning Strategy. doi: 10.1109/iccds60734.2024.10560419

[6] Xiang, Li., Shunpan, Liang., Yu, Lei., Chen, Li., Yulei, Hou., Tengfei, Ma. (2024). Relationship Discovery for Drug Recommendation. arXiv.org, abs/2404.12228 doi: 10.48550/arxiv.2404.12228

[7] B., Sivaiah., Grishma., S, Tharun., A, Shashi. (2024). Drug Recommendation System on Sentiment Analysis of Drug Reviews by Using Machine learning. International Journal For Science Technology And Engineering, 12(3):2019-2023. doi: 10.22214/ijraset.2024.59247

[8] Kornpong, Ariyachaipong., Pattayapon, Senaluang., Pokpong, Songmuang., Rachada, Kongkachandra. (2024). Drug Recommendation based on drug details and Optimization using Bee Algorithm. 1-6. doi: 10.1109/icci60780.2024.10532744

[9] Kadime, Göğebakan., Ramazan, Ulu., Rahib, H., Abiyev., Melike, Şah. (2024). A drug prescription recommendation system based on novel DIAKID ontology and extensive semantic rules. Health information science and systems, 12 doi: 10.1007/s13755-024-00286-7

[10] Sinchana, N, M., Prasanna, Kumar, K, R., Santhosh, B, J. (2023). Model-based Filtering Techniques for Recommendation Systems in Healthcare Domain. 978-983. doi: 10.1109/i-smac58438.2023.10290568

[11] Sawrawit, Chairat., Apichat, Sae-Ang., Kerdkiat, Suvirat., Thammasin, Ingviya., Sitthichok, Chaichulee. (2023). Medication Recommendation using Word Embedding and Recurrent Neural Network. 1-5. doi: 10.1109/bmeicon60347.2023.10321822

[12] Qasem, Kharma., Qusai, Shambour., Abdelrahman, H., Hussein. (2023). A Hybrid Recommendation Model for Drug Selection. Studies in Informatics and Control - ICI Bucharest, doi: 10.24846/v32i3y202307

[13] Junping, Liu., Zhiju, Wan., Xinrong, Hu., Qiang, Zhu. (2024). Safe drug recommendation through forward data imputation and recurrent residual neural network. Applied Soft Computing, doi: 10.1016/j.asoc.2024.111723

[14] Smt., Gousiya, Begum., Pokkuluri, Kiran, Sree. (2023). Drug Recommendations Using a “Reviews and Sentiment Analysis” by a Recurrent Neural Network. Indonesian Journal of Multidisciplinary Science, 2(9):3085-3094. doi: 10.55324/ijoms.v2i9.530

[15] (2023). Adaptive Multi-Hop Deep Learning based Drug Recommendation System with Selective Coverage Mechanism. doi: 10.1109/csnt57126.2023.10134673

[16] C., Silpa., B., Sravani., Dhingra, Vinay., Ch., Mounika. (2023). Drug Recommendation System in Medical Emergencies using Machine Learning. 107-112. doi: 10.1109/ICIDCA56705.2023.10099607

[17] (2023). A drug recommender system Based on Collaborative Filtering for Covid-19 patients. doi: 10.1109/csicc58665.2023.10105347

[18] (2022). RECOMMED: A Comprehensive Pharmaceutical Recommendation System. doi: 10.48550/arxiv.2301.00280

[19] Theresa, Omodunbi., G., E., Alilu., Rhoda, Ikono. (2022). Drug Recommender Systems: A Review of State-of-the-Art Algorithms. 1-8. doi: 10.1109/ITED56637.2022.10051591

[20] Hadi, Al, Mubasher., Ziad, Doughan., Layth, Sliman., Ali, Massoud, Haidar. (2023). A Novel Neural Network-Based Recommender System for Drug Recommendation. Communications in computer and information science, 573-584. doi: 10.1007/978-3-031-34204-2_46

[21] S. Garg, "Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 175-181, doi: 10.1109/Confluence51648.2021.9377188.

[22] Koganti, G., Edupuganti, S., Shaik, A., & Meena, S. D. (2023, October). Drug recommendation system based on analysis of drug reviews using machine learning. In AIP Conference Proceedings (Vol. 2869, No. 1). AIP Publishing.

[23] G. K. Prashanth and H. S. Yashaswini, "Drug Recommendation System Using Machine Learning," 2024 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), Tumakuru, India, 2024, pp. 1-5, doi: 10.1109/ICSSES62373.2024.10561403

[24] V, Akshaya., B, Raja, Ruben., K, Santhosh., M., Nazreen, Banu. (2024). Drug Review: Sentiment Analysis, Using Advanced Machine Learning. doi: 10.1109/ic3iot60841.2024.10550408

[25] A. Jindal, R. Kamboj, S. Pathak, K. Dubey and A. Vajpayee, "Disease Prediction Based on Symptoms and Drug Recommendation," 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2024, pp. 1-6, doi: 10.1109/ICRITO61523.2024.10522185

[26] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. ArXiv, abs/2005.11401.

[27] Jiang, Z., Xu, F.F., Gao, L., Sun, Z., Liu, Q., Dwivedi-Yu, J., Yang, Y., Callan, J., & Neubig, G. (2023). Active Retrieval Augmented Generation. ArXiv, abs/2305.06983.

[28] Wang, C., Ong, J., Wang, C., Ong, H., Cheng, R., & Ong, D. (2023). Potential for GPT Technology to Optimize Future Clinical Decision-Making Using Retrieval-Augmented Generation. Annals of Biomedical Engineering, 1-4.

[29] Christian, Y., Siahaan, M., & Hansvirgo, H. (2024). Designing a Web-Based Light Novel Application with an LLM-Powered Chatbot Recommendation System Using Scrum Methodology. METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi.

[30] Wang, C., Ong, J., Wang, C., Ong, H., Cheng, R., & Ong, D. (2023). Potential for GPT Technology to Optimize Future Clinical Decision-Making Using Retrieval-Augmented Generation. Annals of Biomedical Engineering, 1-4.