

Personalized Drug Recommender System to Aid Medical Professionals

Adya H. S¹, Aryaman. M², Kushal. N. G³, Thejas B.R⁴, Vijay. M Kashyap⁵

Department of Artificial Intelligence and Machine Learning, DSATM, India¹⁻⁵

Abstract: We propose a medical recommender system that provides disease detection and classification along with personalized recommendations for medical requirements based on the user's current diagnosis, medical history, their susceptibility, and genetics. The objective of the system is to assist medical professionals in their decision-making process and used in situations where access to the required or qualified medical professional is unavailable and immediate treatment is to be provided.

Keywords: Personalized Drug Recommendation, Medical Recommender System, Medical Informatics, Machine Learning in Healthcare, Electronic Health Records (EHRs).

I. INTRODUCTION

The advances in technology with a combination of modern medicine have paved a way for the field of Medical Informatics. Medical Informatics applies modern methods like big data collection and its analysis to provide valuable insights about patients and their conditions. Medical Informatics deals with collecting, storing and performing analysis of patient data in order to make better and informed decisions relating to patient treatment, care and wellbeing. It encompasses a broad range of topics, including electronic health records (EHRs), health information exchange (HIE), disease classification and recommendation system, The primary goal is to enhance the quality, safety, and efficiency of healthcare services by creating systems to assist healthcare professionals. Medical Informatics lays the foundation for creation of a personalized drug recommender system as it deals with the process of collection and storage of patient data. IBM's artificial intelligence machine Watson Health [1] is already able to find and recommend a suitable treatment for patients based on other patients' outcome and evidence-based medicine. IBM claims that 81% of individuals in the healthcare domain who are familiar with Watson Health confirmed that it has a positive impact on their business. This demonstrates that using technology and analytics in the field of medicine is beneficial and assists in the overall process. The study aims to lay the foundation for creating a personalized drug recommender system to aid medical professionals in the decision-making process to reduce treatment time in emergencies.

II. RELATED WORK

Kaur, et. al [2] discussed the challenges in the healthcare industry and the emerging field of predictive medicine. The focus is on leveraging predictive models, laboratory testing, and genetic analysis to anticipate disease susceptibility and optimize treatment outcomes. The integration of medical informatics, which combines computer science, information science, and healthcare, plays a crucial role in improving the collection, storage, retrieval, and application of health and biomedical information. Jensen, et.al [3] propose the utilization of electronic health records (EHRs) for medical research and clinical care, emphasizing the potential for establishing patient-stratification principles, uncovering disease correlations, and gaining a finer understanding of genotype-phenotype relationships. However, they acknowledge that ethical, legal, and technical challenges currently hinder the systematic deposition and mining of such data in EHRs. The article discusses the need to overcome these obstacles to unlock the full research potential of clinical data.

Hamon and Grabar [4] propose a system designed for the analysis of narrative clinical documents with the specific objective of extracting medication related information. Recognizing the integral role of pharmacotherapy in medical care, the study aims to address the challenge of buried medication occurrences within narrative texts. The authors employ text mining methods to design a system capable of not only extracting explicit mentions of medications but also deducing information about medications not covered by dictionaries.

Bates, et.al [5] advocate for the transformative potential of big data and clinical analytics in the US healthcare system. With the increasing adoption of electronic health records, they identify six key use cases where leveraging these technologies can notably reduce healthcare costs.

These use cases include managing high-cost patients, minimizing hospital readmissions, optimizing triage processes, early detection of patient decompensation, preventing adverse events, and refining treatment strategies for diseases affecting multiple organ systems. The authors highlight the importance of deriving actionable insights from clinical analytics, specifying the types of data required for analysis, and outlining the essential infrastructure, encompassing analytics tools, algorithms, registries, assessment scores, and monitoring devices. The paper further addresses policy implications related to regulatory oversight, privacy concerns, and the imperative for research support in the realm of analytics. Overall, the proposal underscores the significant opportunities for enhancing care delivery efficiency and cost-effectiveness through the strategic application of big data and clinical analytics in the healthcare landscape. Qian, et.al [6] present an interactive patient risk prediction method, actively engaging medical experts in responding to queries pertaining to the relative similarity of patients. The effectiveness of this approach is empirically demonstrated through rigorous exploration on benchmark and real clinic datasets. Intriguingly, the findings suggest that querying relative similarities not only proves effective inpatient risk prediction but, in some instances, outperforms the conventional practice of seeking absolute answers, thus opening new avenues for enhanced prediction accuracy in medical contexts. Bandyopadhyay, et.al [7] proposes a novel approach to predicting the risk of cardiovascular (CV) events in individual patients by leveraging electronic health data (EHD) from large healthcare systems. While existing risk prediction models are often built from homogeneous and limited-sized epidemiological cohorts, the authors advocate for the utilization of EHD, which offers access to vast, diverse, and contemporary patient populations. The unique challenges posed by EHD, such as missing risk factor information, non-linear relationships, and varying effects in different patient subgroups, necessitate innovative machine learning methodologies. The authors introduce a machine learning approach based on Bayesian networks trained on EHD to forecast the probability of a CV event occurring within a 5-year timeframe. Devarajan, et.al [8] propose an innovative approach to address the challenges associated with diabetes, characterized by a high prevalence and low control leading to increased premature mortality rates. The primary focus is on maintaining optimal blood glucose levels, crucial for realizing significant medical benefits and reducing diabetes-related risks. Recognizing the limitations of real-time continuous monitoring of blood glucose levels, the authors emphasize the importance of considering additional factors such as ECG and physical activities to avoid potential misjudgments in medication. To meet the ever-growing demand for an omnipresent healthcare system, the proposed solution incorporates emerging technologies like the Internet of Things (IoT) and cloud computing. However, the authors acknowledge the computational complexity, high latency, and mobility issues associated with these technologies. To overcome these challenges, they introduce an energy-efficient fog-assisted healthcare system specifically designed for blood glucose level maintenance. The utilization of the J48Graft decision tree enhances the prediction accuracy of diabetes risk levels. Additionally, the integration of fog computing ensures the immediate generation of emergency alerts for precautionary measures. Experimental results presented in the paper demonstrate the improved performance of the proposed system in terms of energy efficiency, prediction accuracy, computational complexity, and latency.

Chen, et.al [9] propose an approach to address the challenges associated with predicting chronic disease outbreaks in disease-frequent communities amidst the exponential growth of big data in biomedical and healthcare domains. Recognizing the limitations posed by incomplete medical data on analysis accuracy, the study focuses on streamlining machine learning algorithms. Drawing on real-life hospital data from central China between 2013 and 2015, the authors experiment with modified prediction models and introduce a latent factor model to reconstruct missing data, overcoming the hurdle of incomplete information. Specifically targeting the regional chronic disease of cerebral infarction, the paper presents the Convolutional Neural Network-based Multimodal Disease Risk Prediction (CNN-MDRP) algorithm. Notably, the proposed algorithm integrates both structured and unstructured data, offering a novel perspective in medical big data analytics. Comparative evaluations showcase the algorithm's impressive 94.8% prediction accuracy, outperforming typical prediction algorithms, and exhibiting a faster convergence speed compared to a CNN-based unimodal disease risk prediction model. The research underscores the importance of a comprehensive and innovative approach to enhance the precision of predicting chronic disease outbreaks in diverse communities. Shimada, et.al [10] proposes the development of a decision support system designed to assist doctors in choosing appropriate first-line drugs. The system's functionality involves classifying patients' abilities to defend themselves against infectious diseases, assigning a risk level for infection. The evaluation of the prototype system demonstrated a correlation between the risk levels determined by the system and the decisions made by specialists. Nilashi, et.al [11] propose a novel knowledge-based system for diseases prediction using a combination of clustering, noise removal, and prediction techniques. The focus is on leveraging data mining techniques, particularly Classification and Regression Trees (CART), to generate fuzzy rules that form the basis of the proposed system. The method is tested on various public medical datasets, including Pima Indian Diabetes, Mesothelioma, WDBC, StatLog, Cleveland, and Parkinson's telemonitoring datasets. The results demonstrate a significant improvement in diseases prediction accuracy, indicating the effectiveness of the proposed approach. The combination of fuzzy rule-based CART with noise removal and clustering techniques is highlighted as particularly beneficial for predicting diseases from real-world medical datasets.

The authors suggest that this knowledge-based system could serve as a valuable clinical analytical method, assisting medical practitioners in healthcare practices by providing enhanced predictive capabilities for diseases based on comprehensive medical datasets. Harimoorthy and Thangavelu [12] proposed a general architecture for predicting multiple diseases in the healthcare industry. This system was experimented to set features of Chronic Kidney Disease, Diabetes and Heart Disease dataset using improved SVM-Radial bias kernel method. Babu, et.al [13] introduces a novel methodology aimed at refining the effectiveness of medical disease prediction systems by integrating advanced techniques such as Grey Wolf Optimization (GWO) and Autoencoder (AE) based Recurrent Neural Network (RNN). The proposed architecture meticulously incorporates key stages, encompassing data acquisition, preprocessing, and the strategic use of GWO for attribute selection, culminating in disease classification facilitated by an AE-based RNN. Noteworthy datasets, including Hungarian, PID, Mammographic Masses, Cleveland, and Switzerland, are employed for comprehensive evaluation. Nayak and Tulasi [14] emphasizes the prevalence of clinical errors in prescription due to limited knowledge and the need for an effective medication selection process. Their paper focuses on developing a medication recommender system using the concept of sentiment analysis and feature extraction to recommend drugs based on patient reviews. They also state that metrics like Linear SVM on TF-IDF outperform other models. John, et.al [15] proposed an advanced medical recommendation system designed to analyze clinical documents and facilitate personalized medication recommendations. Comprising three key modules Extraction, Clustering, and Recommendation. The system employs a Stanford parser for context-based extraction of medical terms from clinical texts. These terms undergo thorough validation against the UMLS database, establishing associations with diseases and forming a symptom list. The Clustering Module utilizes association rule mining and the Kmeans algorithm to identify patterns and cluster clinical documents based on extracted features. Association rule mining is further applied to establish relationships between diseases. The Recommendation Module employs collaborative filtering, comparing document clusters to find similarities between diseases and providing tailored medication recommendations. The proposed system demonstrates promising results in experiments, emphasizing precision, recall, and accuracy metrics, suggesting its potential as a valuable tool in medical diagnosis and treatment recommendation. Stark, et.al [16] in their study focused on construction of a medical recommender system, delineated its development, conducted a comprehensive analysis of distinct systems, and proposed prospective avenues for further investigation. The primary objective of the recommender system is to propose optimal drug options for patients by drawing insights from the success of similar cases. The investigation in this paper focuses on two principal categories of recommendation engines: i). those grounded in machine learning and data mining, and ii). those adopting an ontology and rule-based methodology. The intent is to delve into the intricacies of these approaches and discern their comparative effectiveness. Nguyen, et.al [17] propose an innovative approach to address the challenges in healthcare data classification, specifically focusing on the uncertainty and highdimensional nature of medical data. The proposed method, named GSAM (Genetic Algorithm-based Fuzzy Standard Additive Model), integrates a fuzzy standard additive model with a genetic algorithm. This combination aims to effectively handle uncertainty and computational challenges associated with healthcare data. The GSAM learning process involves three sequential steps: rule initialization through unsupervised learning using adaptive vector quantization clustering, evolutionary rule optimization using genetic algorithm, and parameter tuning through gradient descent supervised learning. Wavelet transformation is employed to extract discriminative features from high-dimensional datasets, contributing to the method's efficiency, especially with a reduced computational burden when deployed with a small number of wavelet features. Vilar, et.al [18] propose a comprehensive review of methodologies and recent advances in using data mining to detect drug-drug interactions (DDIs) that impact patients, with a particular focus on drug safety in both drug development and post marketing pharmacovigilance. The paper highlights the significance of DDIs in causing adverse drug effects, leading to increased risks for patients and elevated costs for public health systems. The primary aim of drug safety researchers is to develop methods for monitoring and discovering potential DDIs that may harm the population.

Naz, et.al [19] propose an ontology-driven system called the Advanced Drug-Drug Interaction (ADDI) system to address the challenges posed by the rapid growth of data in the pharmaceutical domain, specifically focusing on Drug-Drug Interaction (DDI) analysis. The existing tools are noted for lacking automated DDI analysis, and the interaction details are not machine-readable, requiring additional processing by pharmacists for extraction. The ADDI system is designed to assist physicians and pharmacists in identifying DDI effects by providing ontological definitions and semantic relations among various elements related to drug interactions. These elements include diseases, drugs, ingredients, action mechanisms, physiologic effects, dosage formulations, administration methods, DDI mechanisms, DDI types (such as Antagonism, Synergism, Potentiation, and Interaction with metabolism), DDI reactions, their frequency, and duration. The proposed system serves as a Semantic Information Layer (SIL) to address heterogeneity issues and aims to play a significant role in removing barriers for semantic interoperability in the context of DDI analysis. Gottlieb, et.al [20] propose a novel drug-drug interaction (DDI) prediction method called INDI (INferring Drug Interactions). The key innovation of INDI is its capability to infer both pharmacokinetic, CYP-related DDIs (along with their associated metabolizing enzymes, CYPs) and pharmacodynamic, non-CYP associated interactions.

Through cross-validation, INDI demonstrates high specificity and sensitivity levels with an area under the receiver-operating characteristic curve (AUC) exceeding 0.93. When applied to the FDA adverse event reporting system, INDI suggests that 53% of drug events could potentially be linked to either known (41%) or predicted (12%) DDIs. Notably, INDI also predicts the severity level of each DDI upon coadministration of the involved drugs, highlighting the prevalence of severe interactions in clinical practice. Examining regularly taken medications by hospitalized patients, the authors find that 18% of patients receive either known or predicted severely interacting drugs, correlating with a higher frequency of hospitalization.

III. RESEARCH METHODOLOGY

In developing the Personalized Drug Recommendation System, a systematic approach was employed to ensure the reliability and effectiveness of the proposed solution. This section outlines the specific research methods proposed for data collection, preprocessing, feature selection, and the implementation of the machine learning model.

3.1. Data Collection

To construct a comprehensive dataset, electronic health records (EHRs) are to be obtained from various healthcare institutions. These records included patient demographics, medical histories, prescribed medications, and any recorded adverse drug reactions. Additionally, genetic information, where available, can be incorporated to enhance the individualization of the drug recommendation process. The diversity of the dataset is crucial to building a robust and inclusive model.

3.2. Data Preprocessing

The collected data shall undergo a rigorous preprocessing phase to ensure data quality and consistency. The process involves-

- Handling missing values: Employing imputation techniques or removing entries with incomplete information.
- Outlier detection and removal: Identifying and addressing data points that deviate significantly from the norm.
- Standardization and normalization: Scaling numerical features to a standard range for uniformity.

3.3. Feature Selection

The selection of relevant features is pivotal in the success of a personalized drug recommendation system. Through a combination of domain expertise and statistical analysis, key features influencing drug responses should be identified. These features include patient age, gender, medical conditions, genetic markers, and previous drug responses. The feature selection process aimed to maximize the predictive power of the model while minimizing computational complexity.

3.4. Machine Learning Models:

Several machine learning models are being considered and evaluated for their suitability in the drug recommendation system. These models included but were not limited to:

- Collaborative Filtering: Leveraging the collective wisdom of similar patients to make recommendations based on shared characteristics and preferences.
- Content-Based Filtering: Focusing on the specific characteristics of each patient, such as medical history and genetic information, to make personalized recommendations.
- Decision Trees and Random Forest: Exploring decision trees and ensemble methods to capture complex relationships within the dataset.
- Neural Networks: Implementing deep learning techniques, such as neural networks, to extract intricate patterns and correlations for improved accuracy.
- The models will be compared based on their performance metrics during the training and validation phases, with the aim of selecting the most effective approach for personalized drug recommendations.

IV. PROPOSED ARCHITECTURE

The previously discussed methods and implementations focus on recommendation of medicinal requirements based on similarity of their condition to other patients in the past or on sentiments of users of the medicine, these systems cannot assess the disease themselves based on the inputs provided and are unable to provide accurate recommendations; These systems do not focus on providing a personalized system that takes into account the patient's current situation, their medical history, their susceptibility and their genetics, leading to a very generic recommendation not suitable in certain cases. Therefore, we are proposing a recommender system that considers all the factors of a given patient along with their diagnosis to provide an accurate recommendation for every user. the proposed system consists of two parts, i).

A disease classifier and ii). A recommender system. The disease classifier is a machine learning model trained on a classification algorithm to take inputs from the user and classify the disease based on the symptoms provided. The recommender system is used to recommend the most suitable medicinal combination to the patient. The output of the disease classifier feeds into the recommender system as an input along with other initial input features to recommend the best drug for the patient based on their unique requirements. There are three primary reasons for splitting the system into two, i.e. the classifier and the recommender, they are

- i). to create a computationally light system that can run-on low-end hardware/single board computers without issues and
- ii). to help identify alternate diseases, which could be overlooked or not considered, which can be crucial.
- iii). to allow the user to individually access either the recommender or the classifier based on the operational requirements.

The proposed architecture is shown diagrammatically below:

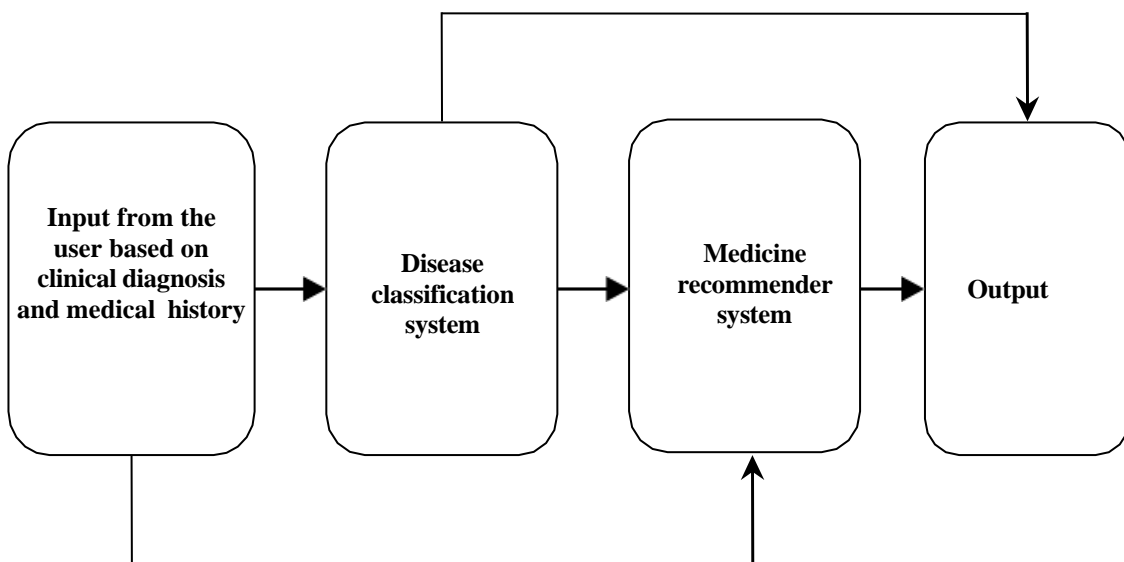


Fig. 1 Proposed system design for the medical recommender system.

V. DISCUSSION

The introduction underscores the significant intersection of technology and modern medicine through Medical Informatics, particularly in the context of developing a Personalized Drug Recommendation System. Notable is the mention of IBM's Watson Health, illustrating the positive influence of artificial intelligence on healthcare decision-making. The related work section provides a comprehensive overview of existing research, addressing challenges in healthcare and the emergence of predictive medicine, aligning well with the study's objective.

The proposed research methodology outlines a systematic approach, encompassing diverse data collection, rigorous preprocessing, and feature selection, along with the evaluation of various machine learning models. The discussion on approaches and the selection of models reflects a meticulous methodology aimed at improving the efficiency and personalization of drug recommendations. The study sets a robust foundation for future research, emphasizing the evolving landscape of healthcare technology and the need for personalized approaches to enhance patient outcomes.

VI. CHALLENGES

A few of the challenges we could face during the development of the proposed system are:

- **Availability of Suitable Data:** The type of data required for the implementation of the proposed system is not publicly available and getting access to authentic patient data is a major challenge faced by us.

- **Data Privacy and Security:** Dealing with electronic health records involves sensitive patient information. Ensuring strict adherence to privacy regulations and implementing robust security measures to protect patient data from unauthorized access is a crucial challenge.
- **Data Quality and Standardization:** Electronic health records may come from various sources, leading to inconsistencies in data formats, terminology, and quality. Standardizing and cleaning the data to ensure accuracy and reliability pose significant challenges.
- **Integration of Genetic Data:** While incorporating genetic information enhances personalization, integrating genetic data seamlessly into the recommendation system is challenging. Genetic data often requires specialized handling, and its interpretation can be complex.
- **User Acceptance and Trust:** Healthcare professionals may be skeptical about relying on automated systems for critical decisions. Building trust and ensuring user acceptance of the recommendation system can be a significant challenge.

VII. CONCLUSION

The proposed system aims to address the evolving landscape of healthcare by integrating technology and modern medicine through the lens of Medical Informatics. The focus on developing a Personalized Drug Recommendation System aims to enhance the efficiency and effectiveness of healthcare decisionmaking, particularly in emergency situations or when access to qualified medical professionals is limited. The study builds upon the foundation laid by existing research in predictive medicine and medical recommender systems. The outlined research methodology systematically addresses key aspects of data collection, preprocessing, feature selection, and the evaluation of machine learning models. The proposed architecture, comprising a disease classifier and a recommender system, seeks to provide a holistic solution that considers individual patient factors such as diagnosis, medical history, susceptibility, and genetics.

This approach sets the system apart by aiming to create a computationally light and user-friendly application capable of running on low-end hardware, while also allowing users to access either the classifier or the recommender independently based on operational requirements. However, the development of such a system is not without its challenges. Issues related to data availability, privacy, security, quality, and standardization pose significant hurdles. The integration of genetic data and the establishment of user trust in automated decision-making processes add additional layers of complexity. Overcoming these challenges will require collaborative efforts from researchers, healthcare professionals, and regulatory bodies to ensure the ethical and responsible implementation of advanced technology in the medical field.

REFERENCES

- [1]. IBM, "IBM Watson Health," 2017. [Online]. Available: <http://www.ibm.com/watson/health>.
- [2]. I. Kaur, . R. Garg, T. Kaur and G. Mathur, "Using Artificial Intelligence to Predict Clinical Requirements in Healthcare," *Journal of Pharmaceutical Negative Results*, 2023.
- [3]. P. B. Jensen, L. J. Jensen and . B. Søren, "Mining electronic health records: towards better research applications and clinical care," *Nature Reviews Genetics*, 2012.
- [4]. T. Hamon and N. Grabar, "Linguistic approach for identification of medication names and related information in clinical narratives," *J Am Med Inform Assoc*, pp. 549-554, 2010.
- [5]. D. B. W, S. Saria, L. Ohno-Machado, A. S. and G. Escobar, "Big Data in Health Care: Using Analytics to Identify and Manage High-Risk and High-Cost Patients," *Health Affairs*, 2014.
- [6]. Q. B, W. X, C. N, L. H and G. J. Y, "A relative similarity based method for interactive patient risk prediction," *Data Mining and*, vol. 29, pp. 1070-1093, 2015.
- [7]. Bandyopadhyay, Sunayan and e. al, "Data mining for censored time-to-event data: a Bayesian network model for predicting cardiovascular risk from electronic health record data.," *Data Mining and Knowledge Discovery*, no. 29, pp. 1033-1069, 2015.
- [8]. D. Malathi, S. V, V. V and R. Logesh, "Fog-assisted personalized healthcare support system for remote patients with diabetes," *Journal of Ambient Intelligence and Humanized Computing*, 2019.
- [9]. C. M, H. Y, H. K, W. L and W. L, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," *IEEE Access*, 2017.
- [10]. K. Shimada, H. Takada, S. Mitsuyama and e. al, "Drug-Recommendation System for Patients with Infectious Diseases," *AMIA Annu Symp Proc*, 2005.



- [11]. M. Nilashi, O. b. Ibrahima, H. Ahmadi and L. Shahmoradi, "An analytical method for diseases prediction using machine learning techniques," *Computers and Chemical Engineering*, pp. 212-223, 2017.
- [12]. K. Harimoorthy and M. Thangavelu, "Multi-disease prediction model using improved SVM-radial bias," *Journal of Ambient Intelligence and Humanized Computing*, 2020.
- [13]. S. B. Babu, S. A, C. G. Babu, .. N. Y. Kumar and K. G, "Medical Disease Prediction using Grey Wolf optimization and Auto," *Periodicals of Engineering and Natural Sciences*, vol. 6, pp. 229240, 2018.
- [14]. L. B. NAYAK and L. N. TULASI, "DRUG Recommendation System Based on Sentiment Analysis of DRUG Reviews Using Machine Learning," *Journal of Engineering Sciences*, vol. 13, no. 12, 2022.
- [15]. A. John, M. Ilyas and V. Vasudevan, "Medication recommendation system based on clinical documents," *International Conference on Information Science (ICIS)*, pp. 180-184., 2016.
- [16]. B. Stark, C. Knahl, M. Aydin and . K. Elish, "A Literature Review on Medicine Recommender," *International Journal of Advanced Computer Science and Applications*, 2019.
- [17]. T. Nguyen, A. Khosravi, D. Creighton and S. Nahavandi, "Classification of healthcare data using genetic fuzzy logic system and wavelets," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2184-2197, 2015.
- [18]. S. Vilar, C. Friedman and G. Hripcsak, "Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media," *Briefings in bioinformatics*, p. 863–877., 2017.
- [19]. T. Naz and M. Akhtar, "Ontology-driven advanced drug-drug interaction," *Computers and Electrical Engineering*, 2020.
- [20]. A. Gottlieb, G. Y. Stein and Y. Oron, "INDI: a computational framework for inferring drug interactions and their associated recommendations," *Molecular Systems Biology*, 2012.