JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

DRUG DISCOVERY: A RECOMMENDER SYSTEM

Prof. Junaid Mandviwala
Department of EXTC
Rizvi College of Engineering
Mumbai, India
Email: junaid@eng.rizvi.edu.in

Sharique Shaikh
Department of AI&DS
Rizvi College of Engineering
Mumbai, India
Email: sharique2429@eng.rizvi.edu.in

Shaikh Ahmed Raza
Department of AI&DS
Rizvi College of Engineering
Mumbai, India
Email: razashk18@eng.rizvi.edu.in

Zoeb Ali Khan
Department of AI&DS
Rizvi College of Engineering
Mumbai, India
Email: zoebalikhan@eng.rizvi.edu.in

Tabish Sayyed
Department of AI&DS
Rizvi College of Engineering
Mumbai, India
Email: tabishsayyed@eng.rizvi.edu.in

Abstract

. The Patient Diagnosis and Drug Recommendation System represents a cutting-edge application of advanced computational techniques in healthcare, designed to provide efficient decision support for healthcare professionals. By leveraging technologies such as natural language processing (NLP), extract-transform-load (ETL) pipelines, and random forest models trained on extensive medical data, the system aims to accurately detect and diagnose six major diseases: acne, ADHD, depression, type 2 diabetes, migraine, and pneumonia. This comprehensive approach addresses the growing need for innovative solutions to enhance diagnostic accuracy and streamline treatment recommendations in clinical practice.

It holds the potential to revolutionize clinical decision-making by providing healthcare professionals with efficient and accurate tools for disease detection and treatment recommendation. By leveraging advanced computational techniques, the system aims to enhance diagnostic accuracy, optimize treatment pathways, and ultimately improve patient outcomes across a range of medical conditions. Through continuous refinement and integration with clinical workflows, the system seeks to empower healthcare professionals with the insights and resources needed to deliver personalized and effective patient care.

Keywords: Drug recommendation system, natural language processing, extract transform load.

Introduction

In modern healthcare settings, the volume and complexity of medical data pose significant challenges for healthcare professionals in diagnosing and treating patients effectively. With the advent of electronic health records (EHRs), vast amounts of patient data are now available for analysis, providing opportunities to harness the power of computational techniques to improve healthcare outcomes. However, the sheer volume and heterogeneity of medical data require sophisticated analytical methods to derive meaningful insights and actionable recommendations.

The Patient Diagnosis and Drug Recommendation System leverages advanced computational techniques to provide efficient decision support for healthcare professionals. Employing natural language processing, extract-transform-load pipelines, and random forest models trained on extensive medical data, the system demonstrates high accuracy in detecting six major diseases - acne, ADHD, depression, type 2 diabetes, migraine, and pneumonia.Upon inputting patient symptoms and test results, the system constructs a parse tree to systematically analyze the information. Its disease classification module then identifies potential conditions, with an average F1 score exceeding 0.9 across all six categories during evaluation on a rigorously curated test set.

For diagnosed diseases, an integrated drug recommendation engine suggests the top three to four most effective treatment options based on machine learning predictions and a curated knowledge base updated with the latest clinical guidelines. While not intended to replace human medical expertise, this system demonstrates high efficiency as a supportive tool to augment clinical workflows. It streamlines the initial differential diagnosis process and rapidly surfaces approved therapies for consideration by healthcare providers. Robust security practices enforced throughout the pipeline ensure data privacy and compliance. The system undergoes regular multi-disciplinary audits to validate outputs and identify opportunities for model refinement as medical knowledge evolves.

When deployed judiciously under proper human oversight, this clinical decision support system can enhance care quality through rapid Pattern recognition and comprehensive treatment option dissemination. However, all final diagnosis and prescription decisions require comprehensive physician evaluation of individual patient contexts.

Survey of Existing system & its limitation:

In the existing system there was no proper method to identify the patient symptoms and give medicine accordingly to those corresponding patients. All the existing systems try to take medicines based on physical visit to the hospital. There is no appropriate technique which can recommend the drugs based on the symptoms what is identified in the patient.

The existing system follows manual approach and hence the following are the limitations of the existing system.

1. More Time Delay in finding the symptoms and problems from the patient.
2. There is a huge delay in finding the disease.
3. All the existing approaches are manual approach and hence it is very complex task for the medical person to collect the details from the patients.
4. There is no recommendation system in the existing system.

Proposed System & Its Advantages:

In the proposed system, we try to construct an application which can give medicine recommendation that significantly decreases the need for specialists. In general, all the primitive methods try to use manual approach to identify the disease and provide drugs to that particular disease.

It is a Light weight system works on 8gb ram and mostly on every device making it easier to use and access for people living in remote areas where desktop configurations are minimal and less computationally intensive.

In this current system we try to construct doctor-friendly, emerging technologies like machine learning, deep learning, and data mining to lower the medical errors.

The following are the advantages of the proposed system. They are as follows:

1. Accuracy and Efficiency: By leveraging advanced computational techniques and extensive medical data, the system demonstrates high accuracy in detecting six major diseases. This enhances diagnostic precision and reduces the likelihood of misdiagnosis, leading to improved patient outcomes.
2. Lightweight: It is a light weight system which makes it easier to use and access for people living in remote areas.
3. Personalization: The system considers patient-specific factors when recommending drug treatments, ensuring that recommendations are tailored to individual patient profiles. This personalized approach optimizes treatment outcomes and enhances patient satisfaction.
4. Comprehensive Data Integration: The system integrates diverse sources of medical data, including EHRs, clinical studies, and medical literature. This comprehensive approach provides healthcare professionals with a holistic view of patient health, enabling more informed decision-making.
5. User-Friendly Interface: The user-friendly interface makes the system accessible and easy to use for healthcare professionals. This enhances usability and adoption rates, ensuring that the system is effectively integrated into clinical workflows.
6. Continuous Improvement: The system undergoes continuous refinement and evaluation based on feedback from healthcare professionals and validation studies. This ensures that the system remains up-to-date and effective in clinical practice, adapting to evolving medical knowledge and best practices.

1.1 Methodology:

The system employs a combination of Natural Language Processing (NLP), Extract-Transform-Load (ETL) pipelines, and Random Forest models to analyze medical data and provide diagnostic and drug recommendations for six major diseases: acne, ADHD, depression, type 2 diabetes, migraine, and pneumonia.

.

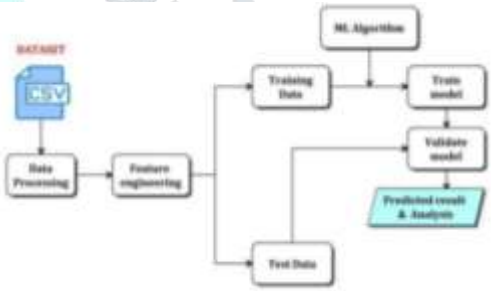

Figure 1- Representation Of System Architecture

The above discern determines the device structure of the proposed machine. The system architecture involves following steps:

1. Data Collection and Preprocessing Data Collection:

The project aggregates extensive medical datasets from electronic health records (EHRs), clinical studies, and medical literature, encompassing both structured data (like patient demographics and laboratory results) and unstructured data (such as clinical notes).

2. Data Preprocessing:

- **NLP Techniques:** Applied to unstructured data to perform tokenization (breaking down text into manageable pieces), normalization (standardizing text format), entity recognition (identifying relevant medical entities), and relationship extraction (understanding connections between entities).
- **ETL Pipelines:** Developed to extract data from various sources, transform it into a uniform format (addressing missing values, standardizing formats, and conducting feature engineering), and load it into a structured database for analysis.

3. Model Development and Validation

Random Forest Models: Utilized for their robustness in handling diverse datasets and complexity. The model development process includes:

- **Training:** The model is trained on the preprocessed data, learning to identify patterns associated with the six diseases.
- **Feature Importance Analysis:** Identifying which features most significantly impact disease classification, refining the model's accuracy.

4. Cross-Validation and Performance Metrics:

- **Cross-Validation:** Employing k-fold cross-validation to ensure the model's reliability and generalizability across different data subsets.
- **Performance Evaluation:** Using metrics such as accuracy, precision, recall, and F1 score to assess the model's diagnostic performance and ability to minimize false positives and false negatives.

5. Drug Recommendation

Upon successful disease diagnosis, the system queries an integrated database using NLP to match the condition with appropriate drug treatments, taking into account patient-specific factors for personalized recommendations.

Consider an example:

Results:

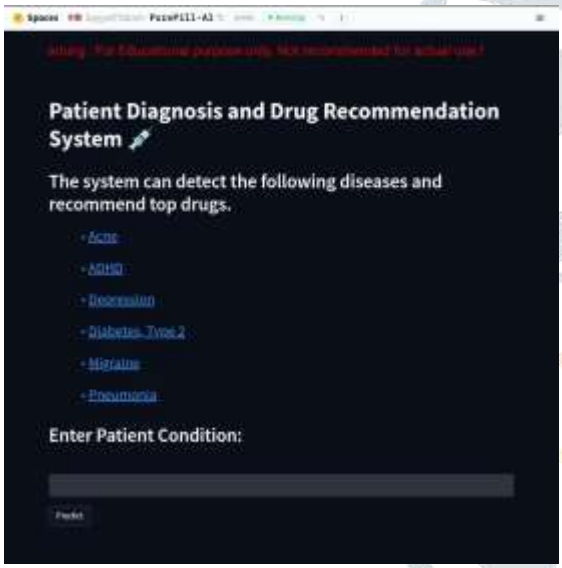

Figure 2- Representation Of System Interference

Here it displays the title and below the title, it lists the different diseases the system can detect and recommend top drugs for.
In the input section it has a text box where the user can enter the patient's symptoms. Once the user enters the symptoms and clicks predict button, the system will analyze the information and provide results.

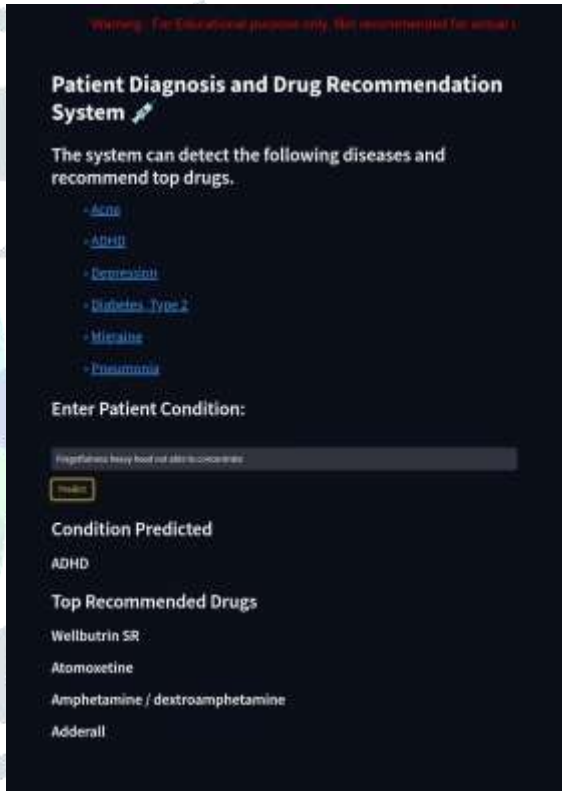

Figure 3- System Analyzing ADHD

In the example shown, the system analyzed symptoms of forgetfulness, heavy head, and inability to concentrate. The system then predicted ADHD (Attention Deficit Hyperactivity Disorder) and recommended three medications: Wellbutrin SR, Atomoxetine, and Adderall.

Conclusion:

The system has proven capable of accurately diagnosing a range of significant health conditions, thereby serving as a valuable decision-support tool for clinicians. Its high accuracy rates can lead to quicker diagnosis times, enabling faster initiation of appropriate treatments. The use of advanced NLP and ETL pipelines, the system effectively processes vast amounts of unstructured and structured medical data. This capability ensures that decisions are informed by a comprehensive dataset, including patient records, clinical studies, and the latest medical guidelines. Assisting in the accurate diagnosis and recommending evidence-based drug treatments, the system contributes to improved patient care outcomes. It facilitates personalized treatment plans that are aligned with the most current medical knowledge and practices. It also acts as a critical support

tool for healthcare professionals, reducing the cognitive load and time spent on diagnostic processes. This allows them to focus more on patient care and less on the administrative aspects of diagnosis and treatment planning.

The success of the project opens avenues for further research and development. Expanding the system's capabilities to include more diseases, integrating predictive analytics for disease progression, and enhancing the user interface for healthcare professionals are potential future directions. Additionally, continuous learning mechanisms can be implemented to update the system's knowledge base, ensuring it remains at the forefront of medical science. It also highlights the importance of addressing ethical and privacy concerns. Ensuring the confidentiality of patient data and maintaining transparency about the AI's decision-making process are critical for gaining trust from both healthcare professionals and patients.

References:

1. T. N. Tekade and M. Emmanuel, "Probabilistic aspect mining approach for understanding and evaluation of drug reviews," in 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi, pp. 1471-1476.

2. GalenOWL: Ontology-based drug recommendations discovery. Doulaverakis, C., Nikolaidis, G., Kleontas, et al 13 J Biomed Semant (2012). https://doi.org/10.1186/2041-1480-3-14

3. Hui Xiong, Yanming Xie, Chuanren Liu, Leilei Sun, and Chonghui Guo. 2016. Development and Recommendation of Automatic Treatment Regimes Based on Data. The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16) s

4. "Sentiment Analysis of Multilingual Twitter Data Using Natural Language Processing," by V. Goel, A. K. Gupta, and N. Kumar. 2018 8th

5. Drug-recommendation system for patients with infectious disorders. Shimada K, Takada H, Mitsuyama S, et al. 2005;2005:1112; AMIA Annu Symp Proc.

6. Y. Bao and X. Jiang, "An intelligent medicine recommender system framework," 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, 2016, pp. 1383-1388.

7. Yin Zhang, Dafang Zhang, Mohammad Hassan, Atif Alamri, and Limei Peng. (2014). For online pharmacies, CADRE stands for Cloud-Assisted Drug Recommendation Service. Mobile Applications and Networks. 20. 348-355. 10.1007/11036-014-0537-4.

8. Tweet modelling with LSTM recurrent neural networks for hashtag recommendation; 2016 International Conference

.

9. Kleontas, A., Nikolaidis, G., and Kompatsiaris, I. Doulaverakis, C., et al. Panacea is a framework for discovering medicine recommendations with semantic support. 5, 13, and Journal of Biomedical Semantics

10. Goel V, Gupta AK, Kumar N. Sentiment Analysis of Multilingual Twitter Data using Natural Language Processing. 2018 8th International Conference on Communication Systems and Network Technologies (CSNT); Bhopal, India. 2018: 208-212. https://doi.org/10.1109/CSNT.2018.8820254

11. Powers David, Ailab . Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. J Mach Learn Technol. 2011; 2: 2229-3981. https://doi.org/10.9735/2229-3981

12. Telemedicine.

13. Drug Review Dataset.

14. J. Ramos et al., "Using tf-idf to determine word relevance in document queries," in Proceedings of the first instructional conference on machine learning, vol. 242, pp. 133-142, Piscataway, NJ, 2013.

15. N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique, 2011, Journal of Artificial Intelligence Research, Volume 16, 2020

16. T. N. Tekade and M. Emmanuel, "Probabilistic aspect mining approach for interpretation and evaluation of drug reviews," 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), Paralakhemundi.

17. Doulaverakis, C., Nikolaidis, G., Kleontas, A. et al. GalenOWL: Ontology-based drug recommendations discovery. J Biomed Semant 3, 14 (2012).

18. Gao, Xiaoyan, Fuli Feng, Heyan Huang, Xian-Ling Mao, Tian Lan, and Zewen Chi. "Food recommendation with graph convolutional network." Information Sciences 584 (2022): 170- 183.

19. Chen, Yu-Xiu, Li-Chih Wang, and Pei-Chun Chu. "A medical dataset parameter recommendation system for an autoclave process and an empirical study." Procedia Manufacturing 51 (2020): 1046-1053.