

# HealthDesk- A One Stop Health Application Using Machine Learning Techniques

Preet Soni  
Computer Engineering,  
Dwarkadas J. Sanghvi  
College of Engineering

Yash Gandhi  
Computer Engineering,  
Dwarkadas J. Sanghvi  
College of Engineering

Vruddhi Mehta  
Computer Engineering,  
Dwarkadas J. Sanghvi  
College of Engineering

Dr. Ramchandra S.  
Mangrulkar, Associate  
Professor, Computer  
Engineering, Dwarkadas J.  
Sanghvi College of  
Engineering

**ABSTRACT** - HealthDesk is a mobile solution for all health-related queries. The application is for all the users of the healthcare system. It covers all the major domains in the hierarchy. People have now started digitizing in this domain too. People have shown great trust in this. This helps us to provide the best. Despite this, individuals often face problems ordering their medicines online. There are high chances of people taking medications without a prescription. So, this application recommends top medicines that are similar to the one being ordered to avoid drug-drug interactions. In the scenario of patient emergency many applications for patient health monitoring and appointment scheduling have been developed. However, in the wake of an emergency, people tend to blank out or are unaware of nearby emergency services. So, the application has a feature that enables users to search the most nearby doctor and provide the doctor with the user's current location. The user can be provided with first aid immediately so that he doesn't succumb to death. The healthcare system has frequent updates. The doctors must remain at par with the updates. However, doctors find it strenuous to sit by and search. The app provides them with relevant news according to their preferences. Concluding, this app covers the most important stakeholders of the healthcare system.

**Keyword** - NLP, DNR, Natural language tool kit, Support vector machine, Machine Learning, SMILES, RDKit.

## 1. INTRODUCTION

In this project the user can call for an emergency whenever needed, the patient who has the application also to help the patient in classifying and suggesting them the best results by making a comparison of data based on x features. The project is categorized into 3 phases. Nowadays, with advancement in technology, there is an increase in number of websites, and mobile applications that are developed to solve real-life problems people were once facing. The scope for the development of this android application is focusing on the healthcare domain, which has three purposes. One is to help people order correct drugs. With users able to buy drugs from the platform, we make a judicious use of data of which drug the patient has ordered. The reason of choosing this is due to increased medical costs, high consultation fees, fear of being put through unnecessary medical tests, many people tend to turn to self-medication. With regard to this they may intake drugs at irregular intervals without realizing it. Increased intakes can cause adverse effects. For example, higher doses of sleeping pills can result in dire consequences. In order to curb the high intake of drug if any, whenever users buy a drug through the app, it alerts the user if the medicines are interacting using Machine learning model based on Clustering technique. The tag line of this application for public users is "Find. Plan. Search." Also, if the user is in any kind of emergency like accident, cardiac arrest, shock, etc., the app will

immediately search for a doctor in your shortest distance and provide the doctor the user's current location. Healthcare is an

ever-changing field of practice with advances in medicine, expanded evidence sources, new treatment options, and changing governmental regulations and models of care. Lifelong learning is not an option in healthcare, it's required by healthcare workers to remain relevant and up to date with such information and continue providing safe, effective patient care. Healthcare professionals are busy round the clock with an increasing amount of patients seeking care. For this reason, developing an app in order to provide the relevant information to healthcare professionals within very less amount of time and in a structured manner is necessary. The app offers immediate up-to-date medical information and reference material at the touch of the finger that is continuously updated.

## 2. LITRATURE SURVEY

### A. Drug Name Recognition:

Approaches and Resources Authors: Shengyu Liu, Buzhou Tang, Qingcai Chen, Xiaolong Wang With increased development in the field of Computer Science, there is an increase in the amount of medical information available. This information can be obtained from various sources like medical journals, medical patent applications, clinical trial documents, medical literature, electronic medical records, etc. In order to exploit the information available, it is necessary to extract valuable information and process it for relevant purposes. Drugs, being a basic entity for medical evaluations become an important target here to be extracted effectively. This Paper mentions an approach termed Drug Name Recognition (DNR), which seeks to recognize drug mentions in unstructured medical texts and classify them into predefined categories, to

help for gaining further insights. Thus, would apply the above approach, which would help us in achieving the objective of providing relevant information to doctors.

The paper mentions basic steps to be followed viz Gather Unannotated Texts from Various Sources, Pre-processing of the text, Selecting Appropriate DNR approach, and finally Post Processing. The main aim would be to select and implement the DNR approach according to the use case. The different approaches cited in the paper include Dictionary based approach, Rule-based approach, Machine Learning approach, and hybrid approach. As of now the Machine Learning Based approach seems to be effective. The process that is mentioned is that this method formalizes DNR as a classification problem or a sequence labelling problem. The tokens generated in pre-processing steps are each designated with a class label following the BIOES-style labelling technique. The four types of drugs (group, nonhuman, brand, drug) are defined. An instance of Bio tagging results of a sentence from the DDI Extraction is demonstrated below. It consists of a sentence and the BIOES tag generated represented as BIOES: Sentence: PXR ligands include a wide variety of pharmaceutical agents, such as antiepileptic drug, Taxol, rifampicin, and human immunodeficiency virus protease inhibitors such as ritonavir and saquinavir. BIOES: PXR-ligands\O include\O a\O wide\O variety\O of\O pharmaceutical\O agentst\O .\such\O as \O antiepileptic \B-group drugs \ I-group ,\O taxol\B-brand \O rifampicin\Bdrug\O and\O human\ B-group immunodeficiency\l-group virus \lgroup protease\l-group inhibitors\I-group such\O ritonavir\B-drug and \O saquinavir \B-drug .\O. This tagging shows us to extract the drug names and so would use it for further implementation to achieve the objective.[1]

### **B. A Counter Medication Classifier Using Machine Learning and Computer Vision Techniques.**

Authors: Eduardo Ceh-Varela , Gandhi HernándezChan , Marisol Villanueva-Escalante, Giner AlorHernández , José Self-medication is one of the major health concerns worldwide and the World Health Organization (WHO) has emphasized correctly

investigating and controlling it. Self-medication or self-prescription can be defined as the action of obtaining and consuming one or more medicines without the advice of a physician nor prescription for treatment. This paper presents the risks about self-medication and proposes a machine learning approach to reduce the risks associated. This paper has the main motive to present a tool that helps in better decision making as to which counter medications to buy. This is done by classifying drugs into different classes/clusters and then providing relevant info like class, description, use of the drug. The authors have used K-means Algorithms to generate clusters. The K factor was determined by comparing two methods, Rule of Thumb and the Elbow Method. Finally, after averaging the K value from both methods, it came out to be 147. Further, the SVM algorithm, a supervised ML algorithm is employed to generate the classification model. The SVM classifier is proposed of achieving an accuracy of 97%. Though the paper implements this along with image processing, the objective was to note the underlying concept of the classification approach used, to aid in this task.[2]

### **C. ANDROID APPLICATION FOR EMERGENCY MEDICAL ASSISTANCE (DOCTORS NEARBY)**

Authors: Dhanesh Sharma, Priyanka Dubey, Navin Singh, Prof. Sulachana Devi The time-lapse in medical help to accident victims is a concern of increasing urgency in India and other countries. To know the locations of victims would be an important function in this case. Google Map API can be used to fetch the location. Objective is to develop a mobile application with integration of this API and other added features. According to statistics , 92.59 % of the total Smartphones sold in 2019 in India were the Smartphones that run on Android Operating System. This paper provides the implementation of an android application in order to provide a solution to the concerned issue . Searching for a nearby doctor during an emergency is one of several mentioned implementations in the application. In case of emergency when the user presses the “emergency button” present in the application, a list of nearby doctors, using the latitude and longitude of the present location and destination location is fetched from the internet using Google Maps API and the distance is calculated. The doctor nearest to the emergency spot, with respect to the minimum distance calculated, is notified about the patient’s location, and the shortest route to the patient’s spot is displayed to the doctor’s screen. This involves implementing different modules. Modules cited in the paper include Login Module: The Signup and Login pages with database connectivity. Emergency Module: The module that covers the functionality of the emergency button which is a major component of the application. Google API module: The display of maps, markers for locations, calculation of routes is handled by this module. Database Module. The User credentials, doctor credentials, which doctor is currently assigned to which patient and all such information is needed to be effectively stored.[3]

### **D. LeMeNo: Personalised News Using Machine Learning**

Authors: D. Khandelwal, D. Shanbhag, A. Shriyan, R. Thorve and Y. Borse The paper demonstrates an application "LeMeNo" that provides user personalized new recommendations. In the day and age, people are not able to keep up with the latest news. With an abundance of news flashing on screens, it is essential to read only the relevant data which appeals to the individual. The need to say "Let Me Know" inspired the author. Based on the user's preferences and engagements in the articles, the news feed is customized. The author presents a literature survey of various topics needed for the recommendation system. He later proposes a model for the same.

The author has further described his implementation steps in detail. Firstly, the author gathers the information using the NLTK library. Keywords are extracted from the titles and descriptions of articles and stored. Then, he classifies them into different categories and subcategories using Naïve Bayesian & SVM algorithms. The application asks for preferences at the user signup. Each article can be liked or disliked. Further recommendations are made, based on this and his engagements. On liking or disliking a particular news article, the system identifies all the keywords associated with that news and creates a vector of these keywords. The author obtains a dataset for a user, which

consists of tags and scores. The score is added by 0.5 for every like. Similarly, subtracted by 0.5 for disliking a particular tag. The author uses this data to match the scores with other users and design a new recommendation based on it. Lastly, the author has created an event timeline to keep the users aware of future events. There is an imperative need to create a solution that can direct users to the relevant articles based on his preferences.[4]

### 3. RESEARCH GAP

The system design has certain drawbacks. Firstly, despite trying to bring the project to the highest accuracy possible. The project can be worked open in the future. With healthcare and medication into the picture, multiple things are to be taken into consideration. We have tried to calculate similarities based on the structure of the drugs. There are various other types of measures to be considered like phenotypic, therapeutic, genomic, etc. Secondly, the role of belief towards IT healthcare. Despite alerting the user on self-medication, if the user is a person who resists technology in terms of reliability, then the application lacks support. The application will not do any good for the user. They may not resist the technology but the effect of it in their lives. Social influence also takes a makes a curve. Another drawback is the use of healthdesk in remote locations. In areas such as villages with limited doctors available already, usage of the app will not make a difference in assistance. During emergencies, it will be difficult to find a doctor nearby.

### 4. PROPOSED ARCHITECTURE WITH MODULAR DESCRIPTION

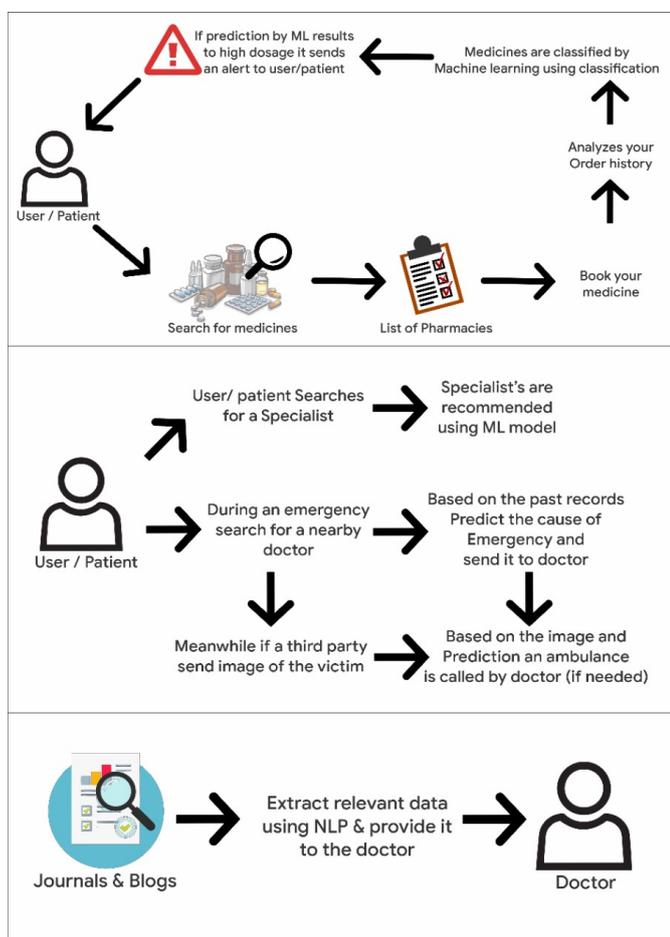


Figure 4.1. Proposed Architecture

The figure 4.1 above is the proposed architecture of HealthDesk. An android application that searches for a doctor during an emergency within the nearest radius from the point of the accident. From a database of hundreds of registered doctors on the application, the nearest doctor is assigned for the patient's assistance, providing the doctor with all the details of the patients through the application including location, medical history, etc. Thus, the doctor can immediately reach out to the patient and initial medication is provided. To curb the issue of high drug intake, an application is developed which keeps a check on the drugs purchased, and alerts the patient if the drug is a Schedule H drug and the patient is unaware of it, or if the drug he purchased from one pharmacy would have a side effect with a drug purchased a few days back from another pharmacy.[5] This is implemented using Machine Learning and applying appropriate learning models on clusters which are different groups of medicines based on similar effects on consumption. This helps prediction to be more dynamic. With the evolvement of information technology, medical documents are available progressively, which contains a magnanimous amount of medical data, such as medical entities and relations between them. It is crucial to extract valuable information from them, to take leverage of medical texts. Drugs being one type of basic medical elements, also need to be recognized. Drug names that appear in unstructured medical texts can be classified into predefined categories. This can be done by a technique termed DNR (Domain Name Recognition), which is a fundamental task of medical information extraction and is a key component of many medical relation extraction systems and applications (e.g., information management, drug discovery, information tracking, information retrieval, clinical decision support, and drug development). This project uses NLP with the abovementioned concept of DNR and thus provide relevant updates to the doctors.[6][7].

### 5. EXPERIMENTATION AND RESULTS

Indians prefer self-medication. They may go to the doctor once, and if they get a similar type of problem, for example, fungal infections, they tend to assume and take the same medicine. Consumption of self-medication can prove to be fatal. In the case of adults, most of them consume multiple drugs, which is called polypharmacy. They might interact and cause unknown problems. Hence the application can guide such people by providing the top medicines which interact with what they want to order. So, there are various ways to find the similarity. The application uses structural similarity. The drug name generates as smiles data, and then we calculate the fingerprint. The project makes use of the Tanimoto coefficient, like cosine similarity, to find the drug similarity.

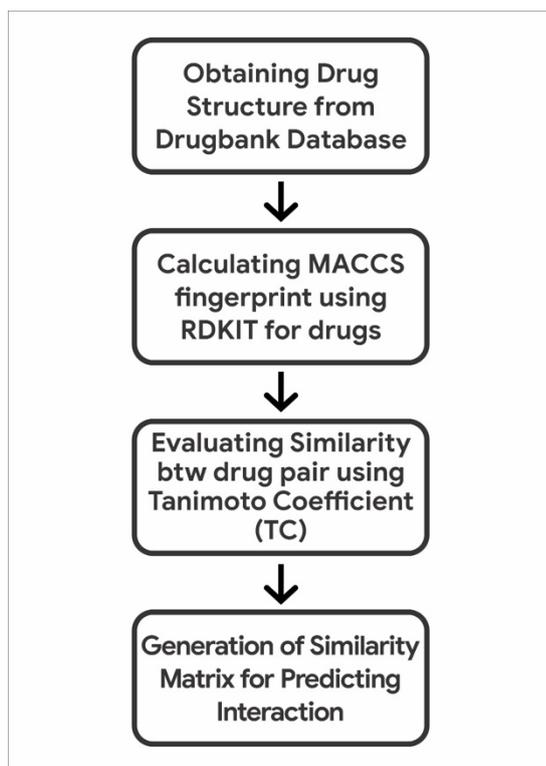


Figure 5.1. Implementation Flow

The model has taken into consideration the Tanimoto coefficient formula for the calculations as in figure 5.2:

$$T = \frac{N_{A\&B}}{N_A + N_B - N_{A\&B}}$$

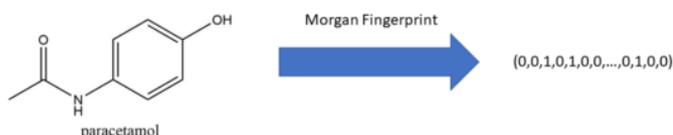


Figure 5.2. Fingerprint

- $N_A$  represents the number of "on" features (bits) in structure A.
- $N_B$  represents the number of "on" features (bits) in structure B.
- $N_{A\&B}$  represents the number of "on" features (bits) common to both fingerprints A and B.

The molecular structure is denoted as strings of special characters using SMILES. The similarity between compounds can be computed, since each molecule is represented as a string.

Example: *Aspirin SMILES:*  
CC(=O)OC1=CC=CC=C1C(O)=O

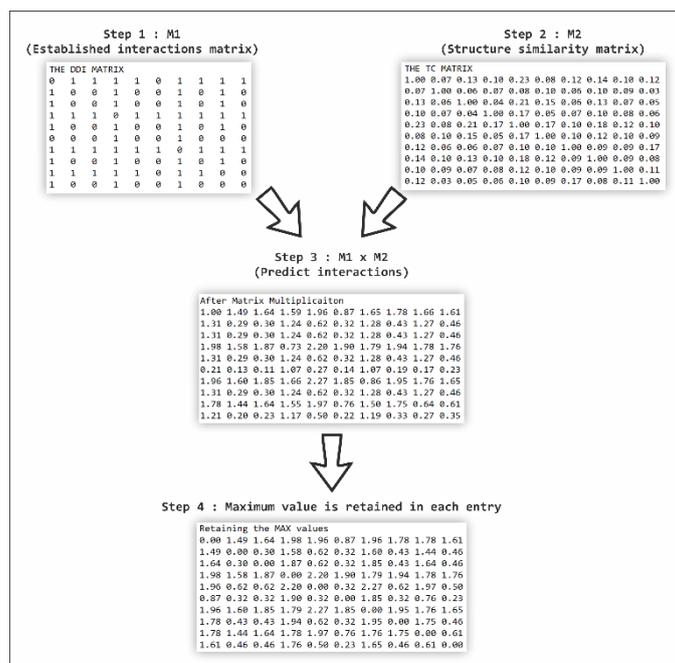


Figure 5.3. Flowchart for DDI

The flow chart in the figure 5.3 above represents establishing a DDI prediction model through a combination. In step one, the M1 is computed by the interaction of approved drugs in the Drug Bank. In step two, M2 is created by the TC values of structural similarity. "1" stands for the interaction which has occurred between two drugs. "0" stands for the opposite situation. Next, multiply the matrix and retain the maximum value in each entry. In step 4, M is formed by retaining the maximum values. Since the interaction of the drug with itself is not considered, set the values in the diagonal of the matrices to "0".[5] Figure 5.4 shows the Confusion Matrix for a TC cut-off value of 0.17 whereas figure 5.5 shows the Confusion Matrix for a TC cut-off value of 0.22.

Accuracy: 0.68  
Mis-Classification: 0.32  
Sensitivity: 0.63  
Specificity: 0.74  
Precision: 0.74  
f\_1 Score: 0.68

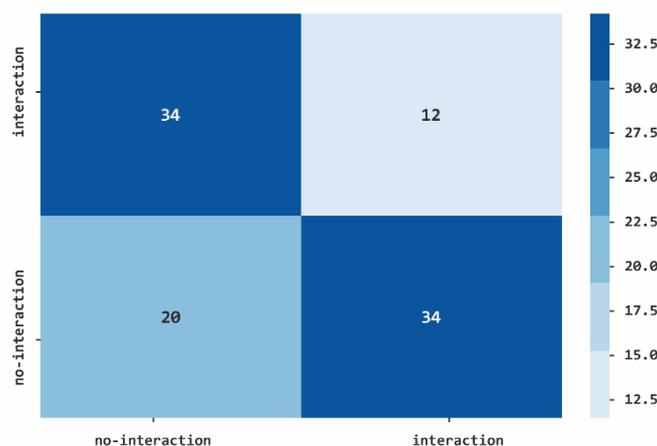


Figure 5.4. Confusion Matrix for threshold value 0.17

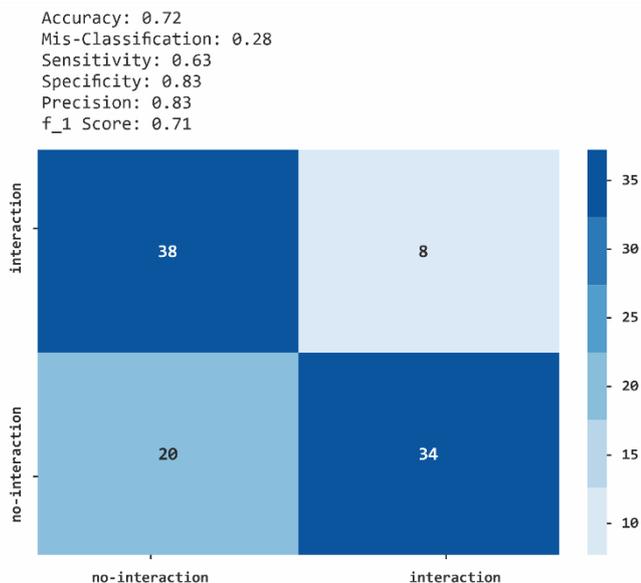


Figure 5.5. Confusion Matrix for threshold value 0.22

According to the results, the performance measures are higher in case when a higher TC cut-off value was taken. An accuracy of 0.68 and 0.72 for the two cut-offs. Since TC value is the measure of the extent of similarity, this model justifies that Drugs with higher degree of similarity would have higher chances of interactions between them

Parameters\ TC	0.22	0.17
Accuracy	0.72	0.68
Mis-Classification	0.28	0.32
Sensitivity	0.63	0.63
Specificity	0.83	0.74
Precision	0.83	0.74
F-1 Score	0.71	0.68

Table 5.1. Comparison for different threshold values.

## 6. CONCLUSION

The project highlighted a Health based application that would help a user/patient to call for an emergency whenever needed. The developed application is a Machine learning-based Health application. The application can be used for calling an emergency through the mobile phone. The user can also search for a particular medicine and a list of pharmacies will be displayed from where the users can directly book the medicine. The application predicts top drugs which interact with the drug being ordered. The last one is designed especially for doctors where the model will extract relevant data using NLP and provide useful information to the doctors through the application.

## 7. REFERENCES

1. Liu, Shengyu & Tang, Buzhou & Chen, Qingcai & Wang, Xiaolong. (2015), "Drug Name Recognition: Approaches and Resources." Information. 6. 790-810. doi : 10.3390/info6040790.
2. Ceh varela, Eduardo & Hernández-Chan, Gandhi & Villanueva-Escalante, Marisol & Alor-Hernández, Giner & Sánchez-Cervantes, José. (2017). A Counter Medication Classifier Using Machine Learning and

- Computer Vision Techniques. Research in Computing Science. 132. 9-22. 10.13053/rcs-132-1-1.
3. Dhanesh Sharma, Priyanka Dubey, Navin Singh, Prof. Sulachana Devi, " ANDROID APPLICATION FOR EMERGENCY MEDICAL ASSISTANCE (DOCTORS NEARBY) ".International Journal of Scientific & Engineering Research, February 2018, ISSN 2229-5518. Web Address : <https://www.ijser.org/researchpaper/android-application-for-emergency-medical-assistance-doctors-nearby.pdf>
4. D. Khandelwal, D. Shanbhag, A. Shriyan, R. Thorve and Y. Borse, "LeMeNo: Personalised News Using Machine Learning," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-4, doi: 10.1109/ICCUBEA.2018.8697560.
5. Huang, De-Shuang & Bevilacqua, Vitoantonio & Premaratne, Prashan. (2016). Intelligent Computing Theories and Application: 12th International Conference, ICIC 2016, Lanzhou, China, August 2-5, 2016, Proceedings, Part I. 10.1007/978-3-319-42291-6.
6. Santiago Vilar, Rave Harpaz, Eugenio Uriarte, Lourdes Santana, Raul Rabadan, Carol Friedman, Drug—drug interaction through molecular structure similarity analysis, Journal of the American Medical Informatics Association, Volume 19, Issue 6, November 2012, Pages 1066–1074, <https://doi.org/10.1136/amiajnl-2012-000935>
7. Vilar S, Uriarte E, Santana L, et al. Similarity-based modeling in large-scale prediction of drug-drug interactions. Nat Protoc. 2014;9(9):2147-2163. doi:10.1038/nprot.2014.151