PATIENT STRATIFICATION BY DISEASE PREDICTION FROM A WIDE-RANGING DATABASE PROGRAM

[1]K. Bhuvaneswari, [2]Dr. S. Babu
[1]Research Scholar, Department of Computer Science and Applications
[2]Assistant Professor, Department of CSA,
Sri Chandrasekharendra Saraswathi Viswa MahaVidyala, Kanchipuram, India

ABSTRACT:
In the era of Web 2.0, questions and answers play an important role to provide an alternative way for knowledge acquisition through the internet. Question retrieval in context to the present community-based question answering (CQA) services does not work well in case of long and complex queries, like the questions. Some of the reasons for it are the verboseness in natural language queries and also the word mismatch that occurs in the queries and the candidate questions in the CQA archive in the process of retrieval. Question retrieval is commonly used in hospital management system as a prediction tool so as to classify the risks/diseases. It aims at delivering a quick and direct solution to patients that automatically produces the respective health advisory which includes name of the disease and the risks associated with the same. In order to analyze the risk factor an additional medical related database must be used (i.e., a dataset that holds 3000 Entries on the various symptoms and its related data). Once the patient enters the required details about the sickness, the duration of the Symptom existing, the Frequency of the sickness and the affected place etc. The system will then immediately analyze the details and begins to classify the problem and its risks. When the risk exceeds the quoted threshold, i.e., when the case is observed to be critical and needs immediate medical attention, then a list of available doctors concerned to the illness is broadcasted to the patient in order to fix appointments. In some cases, the Doctor might recommend some tests and diagnosis to be done in the laboratory. The laboratory feature/module permits the patient to see the recommended tests that are suggested by the doctors. The lab results will be communicated to both the patient and the doctor by lab personnel using this functionality.

Keywords: Disease Prediction, Support vector machine, classification.

INTRODUCTION

All these years the huge amount of Electronic Health Records (EHRs) collected have contributed to a rich database for risk analysis. An EHR holds digitally stored healthcare information about an individual that covers details about observations, diagnostic reports, laboratory tests, medications, patient identifying information, procedures and allergies. A special type of EHR is called as the Health Examination Records (HER) that holds information on the annual general health check-ups. The proposed work in this paper aims at delivering a quick and accessible direct solution to patients that can automatically produce the respective health advisory including the name of the particular disease and the risks associated with the same. A medical related database is being used in this paper to analyze the risk factor (i.e., a dataset with 3000 Entries on symptoms and its associated data). Once the patient enters the required details about the

sickness, the duration of the Symptom existing, the Frequency of the sickness and the affected place etc. The system will then immediately analyze the details and begins to classify the problem and its risks. When the risk exceeds the quoted threshold, i.e., when the case is observed to be critical and needs immediate medical attention, then a list of available doctors concerned to the illness is broadcasted to the patient in order to fix appointments. In some cases, the Doctor might recommend some tests and diagnosis to be done in the laboratory. The laboratory feature/module permits the patient to see the recommended tests that are suggested by the doctors. The lab results will be communicated to both the patient and the doctor by lab personnel using this functionality.

LITERATURE REVIEWS

This method deals with the privacy preserving patient centric clinical decision support system using the naive Bayesian classifier. Using the technology of cloud computing, the processing unit can use any huge medical dataset that can be stored in cloud platform in order to train the naive Bayesian classifier. Once the system is trained it can be applied to diagnose the disease without compromising on the privacy of the data provider. Also, the patient can securely access the top-k diagnosis results on their own preferences in our system. As all the data are encrypted it is ensured that the patient-centric diagnose result retrieval is performed in a privacy preserving way. [1]

At first, the health examination records are converted into graphs and then it associates all the relevant cases together. This approach works well in case of modeling abnormal results that are usually sparse. Next, a multi-typed relationship of the data items is analyzed and then naturally mapped into a heterogeneous graph. In particular the health examination items are denoted using a different type of node in the graph. This enables our method to develop the underlying heterogeneous sub graph structures of individual classes so as to achieve higher performance. Finally these acquired features can be weighted in their own type using a label propagation process on a heterogeneous graph. These in-class weighted features are then used for an effective classification in an iterative convergence process. [2]

This is another contribution to the community based health services. It is a sparsely connected deep learning scheme that can decide on the possible diseases given the questions of health seekers. This scheme is designed using alternative signature mining and pre-training in an incremental way. It supports unsupervised feature learning that includes other wide range of disease types. Therefore, this approach can be generalized and it is scalable when compared to the other previous disease inference that uses shallow learning approaches that are trained on hospital generated patient records with structured fields. The classical deep learning architectures are thickly connected and the number of nodes in each hidden layers are tiresomely adjusted. In contrast, our proposed model is sparsely connected along with improved learning efficiency, and the number of hidden nodes is determined automatically. [3]

An Efficient K-means Clustering algorithm has been implemented to cluster the PHR into many partitions. The EKMC algorithm uses less time than the traditional K-Means Clustering. The proposed DAD algorithm is used to minimize the cost of cloud storage to a greater extent as we are dealing with huge records. The privacy of the patient's PHR is protected through various data anonymization techniques. [4]

The searchable encryption (SE) scheme is a method to incorporate security protection and favorable operability functions together in a system. This approach plays an important role in the e-health record system along with a novel cryptographic primitive named as conjunctive keyword search

with designated tester and timing enabled proxy re-encryption function (Re-dtPECK), that is a kind of a time-dependent SE scheme. This helps patients to delegate partial access rights to others to operate search functions over their records for a limited period of time. The length of the time period for the delegate to search and decrypt the delegator's encrypted documents can be controlled. Moreover, the delegate can automatically deprive the access and search authority after a specified period of time. [5]

OUR CONTRIBUTION

This paper involves a hospital management system that includes a prediction tool to classifying the risks/diseases. . It aims at delivering a quick and direct solution to patients that automatically produces the respective health advisory which includes name of the disease and the risks associated with the same. In order to analyze the risk factor an additional medical related database must be used (i.e., a dataset that holds 3000 Entries on the various symptoms and its related data). Once the patient enters the required details about the sickness, the duration of the Symptom existing, the Frequency of the sickness and the affected place etc. The system will then immediately analyze the details and begins to classify the problem and its risks. When the risk exceeds the quoted threshold, i.e., when the case is observed to be critical and needs immediate medical attention, then a list of available doctors concerned to the illness is broadcasted to the patient in order to fix appointments. In some cases, the Doctor might recommend some tests and diagnosis to be done in the laboratory. The laboratory feature/module permits the patient to see the recommended tests that are suggested by the doctors. The lab results will be communicated to both the patient and the doctor by lab personnel using this functionality.

The main achievements or functionalities in this paper are:

A new efficient iterative algorithm is structured based on Health and modified SVM algorithms to generate more accurate recommendations.

This personalized healthcare model aims at minimizing the Re-admission rates in the hospitals and thereby enabling the cost savings for both healthcare provider and the patient. Usage of large dataset (around 3000) –is indeed a challenging task.

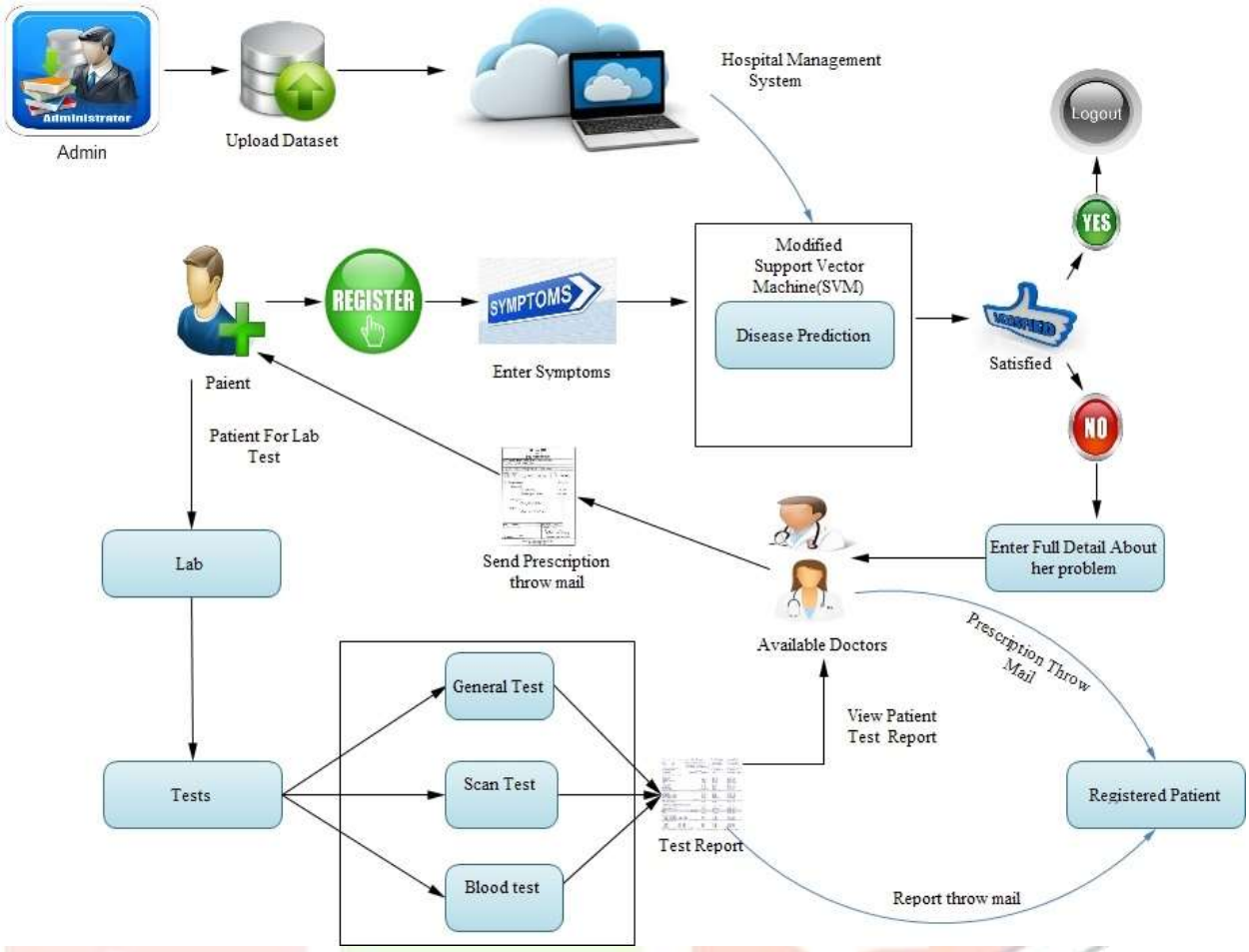


Fig.1.Shows system architecture

Methodology

Proposed Solution

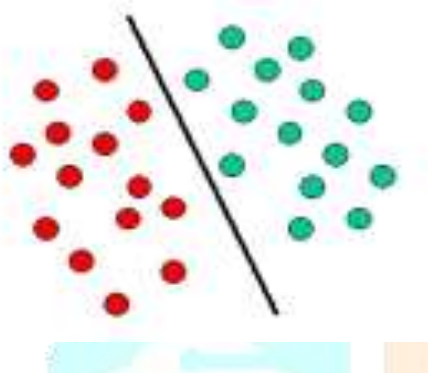
1. In cases of high dimensional space, the support vector machines are highly effective.
2. This method is found to be effective even when the number of dimensions is larger than the number of samples.
3. It is memory efficient because it uses subset of training points as decisive factors for classification.
4. The main advantage of MSVM Algorithm is that if any error occurs in the midst of the execution it can skip that particular step and execute the remaining steps and still produce the accurate output.
5. The speed of the working efficiency is unaltered with the amount of data (large or small).

Modified Support Vector machine

MSVM- Modified Support Vector Machines is used for the classification of data. It is said to be better than the Neural Networks system since the latter method gives unsatisfactory results at instances. The SVM method needs training and testing set of data. Each instance in the training set will contain one target and several attributes. MSVM will help to depict the target value of a data in the testing set with only attributes.

MSVM actually helps to tune the system work correctly through Supervised Learning. When the

data is labeled correctly it helps the system to perform right. In MSVM classification, identification is an important process which is intimately connected to the known classes. This is called as the feature extraction or feature selection. MSVM classification and feature selection can help in identifying key sets that are involved in processes that distinguish the classes.



The above is a classic example of a linear classifier, i.e., a classifier that separates a set of objects into their respective groups (GREEN and RED in this case) with a line. Most classification tasks, however, are not that simple, and often more complex structures are needed in order to make an optimal separation, i.e., correctly classify new objects (test cases) on the basis of the examples that are available (train cases). This situation is depicted in the illustration below. Compared to the previous schematic, it is clear that a full separation of the GREEN and RED objects would require a curve (which is more complex than a line). Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyperplane classifiers. Support Vector Machines are particularly suited to handle such tasks.

Applications of Modified SVM

A modified SVM has proved to be successful in case of pattern classification problems. When Modified Support Vector approach is applied to a particular practical problem it involves resolving a number of questions that are based on the problem definition and the structure involved in it. One of the major challenges in these cases is to choose an appropriate kernel for the given particular application. Though there are standard methods like the Gaussian or polynomial kernel, these approaches fail or to be ineffective if the inputs are discrete structures where more elaborate kernels are in need. Through implicitly defining a feature space, the kernel provides the description language used by the machine to view the data. Once the optimization criterion and the kernel are made to be the key components of the system let's look over some examples.

Text categorization is the process of classifying natural text documents into a fixed number of predefined categories based on the content it holds. In real scenarios a document can be classified to more than one category leading to a multi-class classification problem that can be viewed as series of binary classification problems one for each category. The most ideal solution for text classification in order to retrieve information is the ideal feature mapping for constructing a Mercer kernel. The kernels can incorporate a similarity measure between instances, and it is also expected that the experts working in the specific application domain must have already identified some valid similarity measures.

The traditional classification methods perform poorly when applied on practical scenarios due to the high dimensionality of data. In such cases the Modified Support Vector Machines can eliminate the problems caused by very high dimensional representations. A similar approach can be used for the task of image classification as well. In such case the linear hard margin machines are frequently able to generalize it well. The first real-world task for Modified Support Vector Machines was to be tested on the problem of hand-written character recognition. It is interesting to compare SVMs with other classifiers, and also with other different

Modified SVMs. The comparison shows that they approximately possess the same performance level and furthermore to share most of their support vectors, independently of the chosen kernel. The fact to be accepted here is that Modified SVM can perform the same as these systems that does not require any detailed prior knowledge.

RESULT

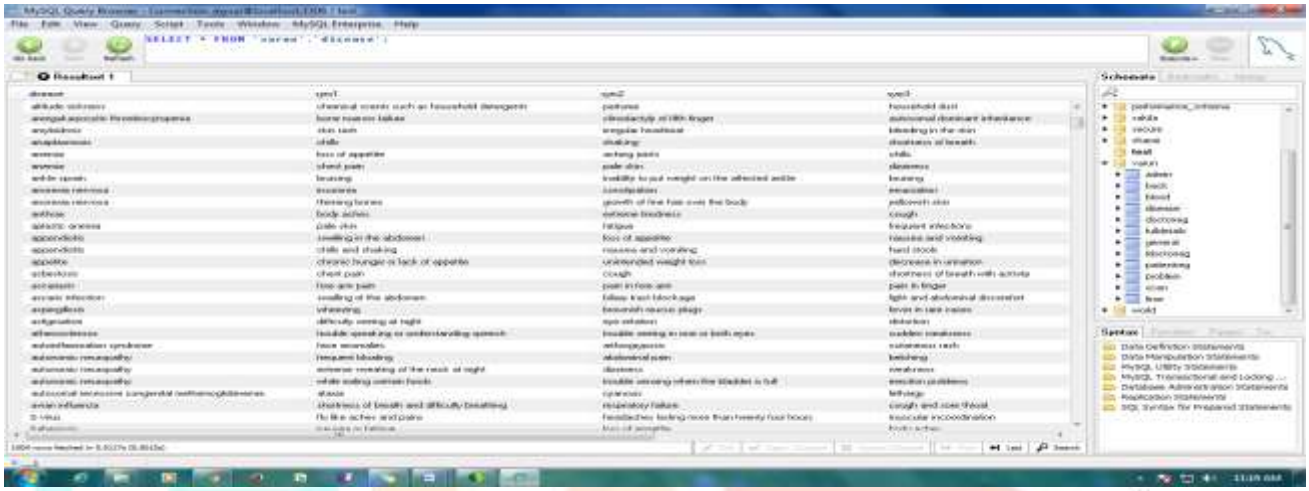


Fig.2.Shows Disease Dataset Table

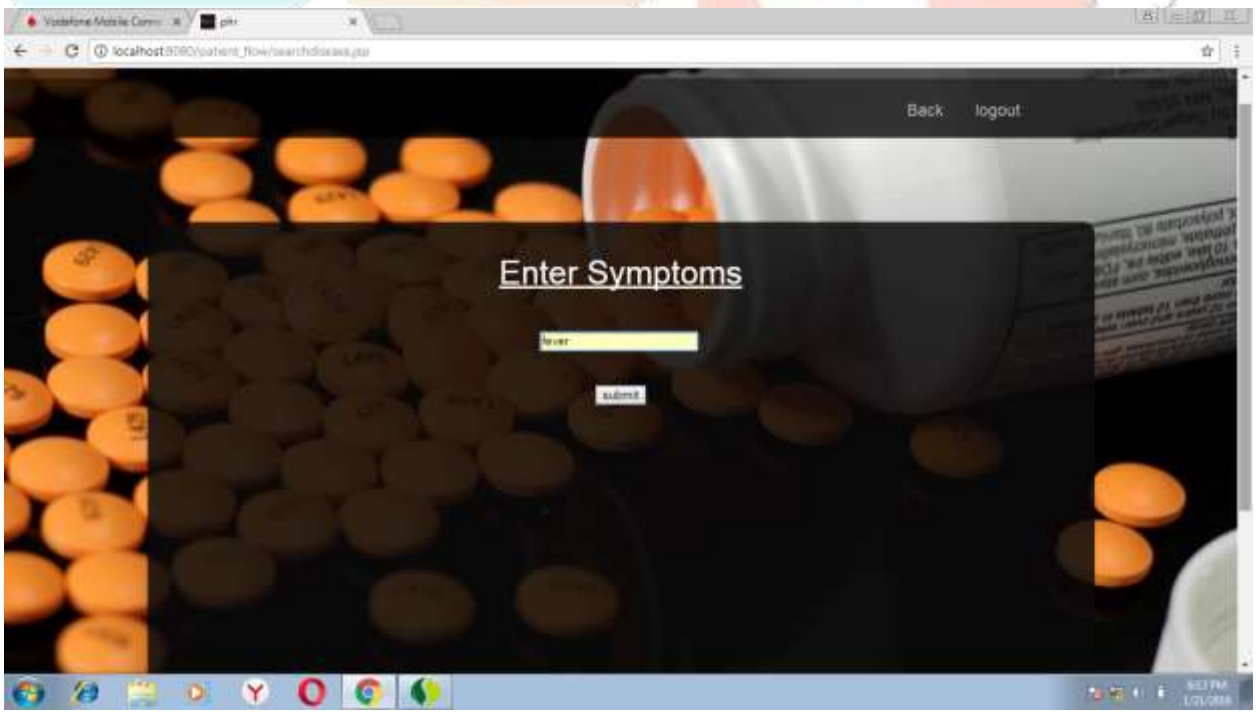


Fig.3 .Shows Search Disease

Fig.4.Disease prediction

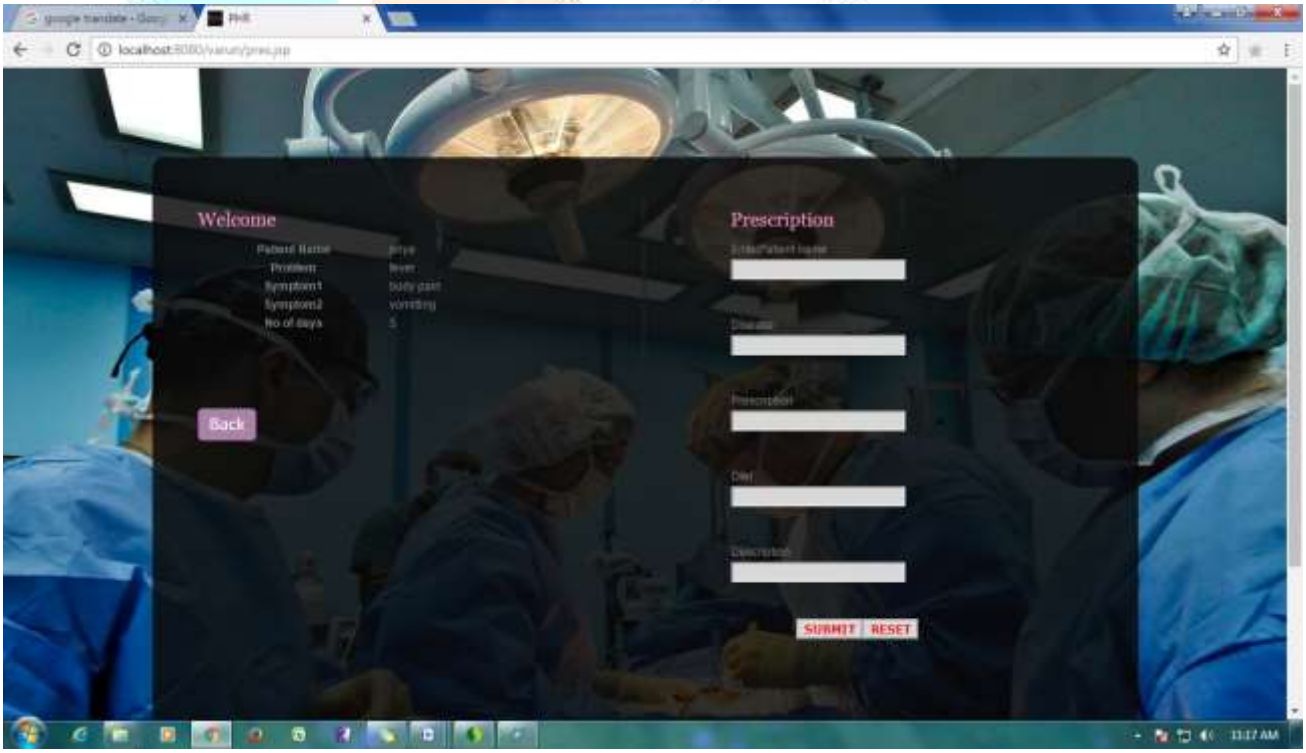


Fig.5.Send prescription

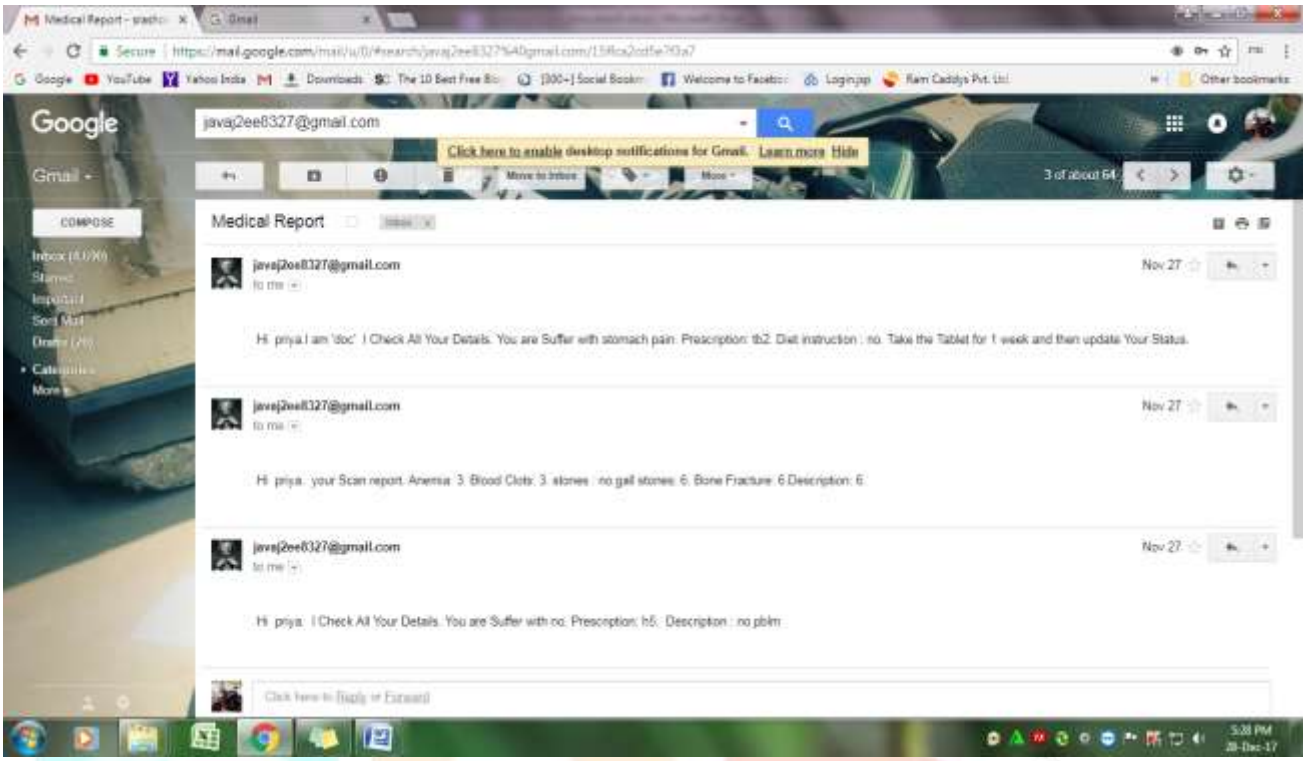


Fig.5.Shows Smtp mail

CONCLUSION

We implement a question retrieval method in hospital management system that is embedded with a prediction tool to classifying the risks/diseases. This approach focuses on delivering a faster and a direct solution to patients. In other words, if patients provide their symptoms of the illness along with other related factors then the model can automatically fetch the respective health advisory that includes the name of the disease and the associated risk. In addition it also aims to design a model provides a personalized healthcare solution. Our proposed framework is architecture for online applications that have computation or memory limitations.

REFERENCES

[1]Ximeng Liu, Rongxing Lu, Jianfeng Ma, Le Chen, and Baodong Qin "Privacy-Preserving Patient-Centric Clinical Decision Support System on Naive Bayesian Classification" DECEMBER 2014

[2] Ling Chen, Xue Li, Quan Z. Sheng, Wen-ChihPeng, John Bennett, Hsiao-Yun Hu, and Nicole Huang"Mining Health Examination Records A Graph-based Approach"

[3]LiqiangNie, Meng Wang, Luming Zhang, Shuicheng Yan, Bo Zhang, and Tat-Seng Chua," Disease Inference from Health-Related Questions via Sparse Deep Learning"
AUGUST 2015

[4] G.Logeswari1, D.Sangeetha2, V.Vaidehi3 "A Cost Effective Clustering based Anonymization Approach for Storing PHR's in Cloud"

[5] Yang Yang and Maode Ma "Conjunctive Keyword Search With Designated Tester and Timing Enabled Proxy Re-Encryption Function for E-Health Clouds" APRIL 2016

[6]Jitendra Madarkar "Security issues of Patient Health Records in E-Hospital Management in Cloud"June 2014.

[7]Nishitha Ramakrishnan1, Sreerekha "Enhancing Security of Personal Health Records in Cloud Computing by Encryption "

[8]Arpana Mahajan, YaskPatel,"Enhancing PHR services in cloud computing: Patient-centric and fine grained data access using ABE" December 2012

[9] Chia-Hui Liu, Tzer-Long Chen, Han-Yu Lin, Fong-Qi Lin, Chih-Ming Liu, En-Ping Wu, Yu-Fang Chung, and Tzer-Shyong Chen "Secure PHR Access Control Scheme in Cloud Computing" May 2013

[10] Joseph A. Akinyele_, Christoph U. Lehmanny, Matthew D. Green_, Matthew W. Pagano, Zachary N. J. Petersonz, Aviel D. Rubin, Johns Hopkins University_ Johns Hopkins Medical Institutions Naval Postgraduate School "Self-Protecting Electronic Medical Records Using Attribute-Based Encryption"