ISSN No:-2456-2165

Computerized Healthcare System Embedded with Machine Learning

Amal S
Dept of Computer Science and Engineering
Mar Athanasius College of Engineering
Ernakulam, India

Athira Lagi
Dept of Computer Science and Engineering
Mar Athanasius College of Engineering
Ernakulam, India

Thomas Kuruvilla
Dept of Computer Science and Engineering
Mar Athanasius College of Engineering
Ernakulam, India

Leya Elizabeth Sunny Dept of Computer Science and Engineering
Mar Athanasius College of Engineering
Ernakulam, India

Abstract:- The project builds a complete computerized health- care system embedded with machine learning applications. A patient's health data are generally stored in a centralized third party, such as the hospital database or cloud, and make users lose control of their health data, which can easily result in privacy leakage. So here we propose a healthcare system which is focused on the complete computerization of healthcare with the help of a unique Medical Card available to every person. Electronic Medical Records(EMRs) are used for storing medical diagnosis and prescription details of patients. The EMRs of a patient lets doctors know about the complete medical history of their patients. This Health Chain can be integrated with machine learning for applications like disease prediction, doctor recommendation. Disease prediction is all about predicting diseases based on earlier symptoms of a patient while doctor recommendation finds out the best and nearby doctors a person should consult based on his/her disease.The system also has a COVID'19 chatbot which interviews the user and predicts his/her probability of being affected with the disease and suggest necessary actions to take.

Keywords:- *Decision Tree, Sentimental Analysis, Naive Bayes Algorithm, K-fold Cross Validation.*

I. INTRODUCTION

Over the past few decades, our medical knowledge has increased. More investigative and treatment options are avail- able; as a result our patients are living longer and we are deal- ing with more chronic conditions. Yet we have not developed a complete computerized healthcare system by integrating all medical providers including hospitals, doctors, laboratories under a single unit.

In this paper, we propose Health Chain - a computer- ized healthcare system which offers a Unique Medical Card based privacy preserving scheme for health data, integrated with machine learning applications. Patient has full control over his/her medical data through Unique Medical Cards. It provides primitives like data integrity and prevents medical frauds. Each patient is identified by a unique medical card. When a person makes use of a medical facility, the summary of the checkup will be appended to his Electronic Medical Records(EMRs) using the unique medical card. These details can be used by doctors to know about the complete medical history of his patient and the medicines he has taken till now. According to the 2014 National Physician Survey [1] 75% of the physicians were using EMRs[10]. Of those, 65% indicated that patient care improved and less than 5% indicated a negative effect on the quality of care they provided.

The value of machine learning in healthcare is its ability to process huge dataset beyond the scope of human capability, and then reliably convert analysis of that data into clinical insights that aid physicians in planning and providing care, ultimately leading to better outcomes, lower costs of care, and increased patient satisfaction. The machine learning embedded in the system predicts diseases based on the symptoms a patient has. Based on the predicted disease, medicines that need to be taken can also be predicted. In addition to all the above features, the system also has a doctor recommendation functionality which uses patients reviews and ratings on doc- tors to recommend best possible doctors available nearby as per users requirements.

II. EXISTING SYSTEM

In the present existing systems there is no integration between various medical providers - hospitals, doctors and laboratories and there is no central authority for coordination between these entities. Electronic Medical Records(EMRs) are generally stored in a centralized third party, namely the hospital database or cloud. This centralized approach has its own drawbacks. The main downside is associated with the concern that a patient doesn't have the ownership of his medical record. Moreover, one cannot access his healthcare records and has no idea about how it is getting used by the centralized third party. There are also concerns where multiple vendors hold different versions of the same patient record that are not validated, resulting in various errors, inconsistencies and incompleteness. Add to that reports of security breaches, tampering of personal data and the ever-present hacking threat. There is lack of self-diagnosis systems for disease prediction or medical suggestions in

ISSN No:-2456-2165

the present healthcare system. An- other matter of notice is that presently one has no knowledge about the treatment qualities offered by doctors in different hospitals, whom should be consulted for a particular disease and so on.

III. SYSTEM ARCHITECTURE

Initially, a centralized authority will verify and register all the authorised medical providers like hospitals, laboratories and provide them with their credentials. Hospitals can login using these credentials and they can register verified doctors. Doctors, in turn, can register patients by providing important patient demographics. Fig. 1 represents the major stakeholders in the healthcare system.

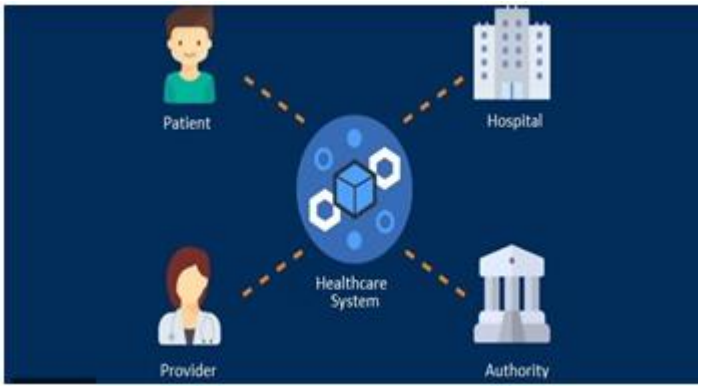

Fig 1:- Major actors in healthcare system

Doctors can manage their profile, add their qualifications and experience and can communicate with other doctors. Doctors also a have provision to post medical articles and health tips which can be viewed by patients and other doctors. Patients follow doctors inorder to see their posts. They can give likes and comments to posts seen on their timeline.

Patients login to their profile to view their medical profile and history. They can set access rights to their medical records and can determine who all can view it. When a patient goes to a doctor, the doctor will scan the QR code present in the Unique Medical Card of the patient to see his profile and medical history. Doctor will add prescriptions and other relevant information during the analysis of that patient to his medical record. The patients can add reviews and rate the treatment quality of doctors. This ratings are used in machine learning to suggest best doctors nearby.

Apart from hospitals and clinics, the system also combines the laboratory unit, so that the patients can obtain their laboratory results directly from their profile, without waiting for the results.

The two most important features of the proposed healthcare system are symptom checker and doctor recommendation system. Symptom checker is used to predict diseases from symptoms entered by patient. This feature is available to patients when they log in. Doctor recommendation is also available when a patient logs in. Here, the user either input the required category or set of symptoms he has and the preferred location. The output will be the best rated doctors of that category, who are nearby.

Another important feature made available to users is the COVID-19 chatbot. The bot interviews users with a set of questions and provides a conclusion regarding the probability of the person getting affected by COVID-19. The questions

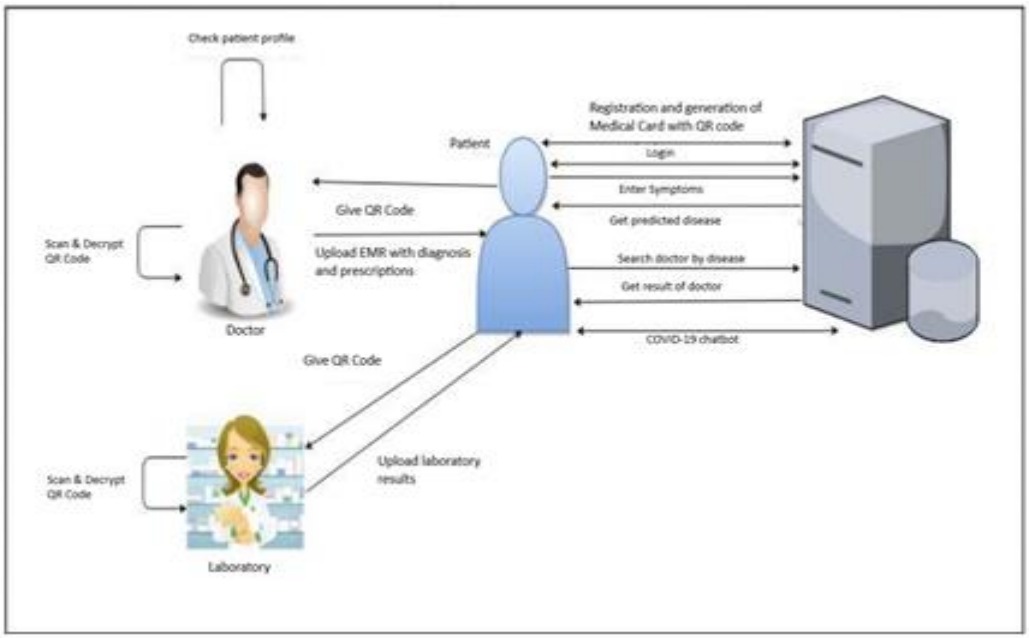

Fig 2:- System Architecture of Healthcare System

ISSN No:-2456-2165

IV. DESIGN

A. Unique Medical Card

Every patient registered in the healthcare system is provided with a Medical Card similar to the Aadhaar Card implemented by UIDAI. It is assumed that the Unique Medical Card is implemented by a central authority, possibly Indian Medical Association. Medical Card contains a patient's basic demo- graphics along with a QR code. The medical providers like hospitals, doctors and laboratories need to scan this QR code using our Healthcare System to view the patient's medical profile and past Electronic Medical Records(EMRs).

Fig 3:- Unique Medical Card

B. Electronic Medical Records

When it comes to patient data management, there are two main issues in the healthcare industry. First, each patient is unique therefore there is no such thing as a common disease or common treatment strategy. What works on a patient might not work on the other due to inter-individual variability. Hence, access to complete medical records is essential in order to adapt the treatment and provide personalised care.Second, in the present existing system, the individuals have no access to their medical records and they cannot decide who has access to their personal medical data.

The solution is Electronic Medical Record (EMR) [2] which provides the complete medical history of an individual in digital form. Whenever a patient consults a doctor through our healthcare system the doctor can view the patient's past med- ical history and treat accordingly. Also a new EMR is created with the summary of consultation along with information like symptoms observed, disease diagnosed, procedures performed and medicines prescribed.

The proposed method also offers every individual the com- plete control over their medical records. They can determine the level of abstraction regarding who all can access it. There are three different levels of abstraction:- Public,Private and Custom. Using this facility, the complete health information of a particular individual is available at a single point, regardless of which all doctors he has consulted.

C. Disease Prediction

Due to big data progress in biomedical and healthcare communities, accurate study of medical data benefits early disease recognition, patient care and community services. When the quality of medical data is incomplete the exactness of study is reduced. Medical facilities need to be advanced so that better decisions for patient diagnosis and treatment options can be made. Machine learning in healthcare aids the humans to process huge and complex medical datasets and then analyze them into clinical insights. This then can further be used by physicians in providing medical care. Hence machine learning when implemented in healthcare can lead to increased patient satisfaction. Here, we try to implement func- tionalities of machine learning in healthcare in a single system. Instead of diagnosis, when a disease prediction is implemented using certain machine learning predictive algorithms then healthcare can be made smart. Some cases can occur when early diagnosis of a disease is not within reach. Hence disease prediction can be effectively implemented. As widely said "Prevention is better than cure", prediction of diseases and epidemic outbreak would lead to an early prevention of an occurrence of a disease. We propose a system which predicts disease given the symptoms using machine learning.

The proposed system bid a machine learning decision tree map, Naive Bayes, random forest algorithm by using structured and unstructured data from hospitals[4]. The dataset values are classified into training and testing sets for the model evaluation procedures. The right model is identified using the k-fold cross validation method[8]. The algorithm without underfitting and overfitting problems is chosen as the machine learning model, which was found out to be decision tree[9].

Model evaluation procedures:

- *Training and testing on the same data:* Rewards overly complex models that ”overfit” the training data and won’t necessarily generalize.

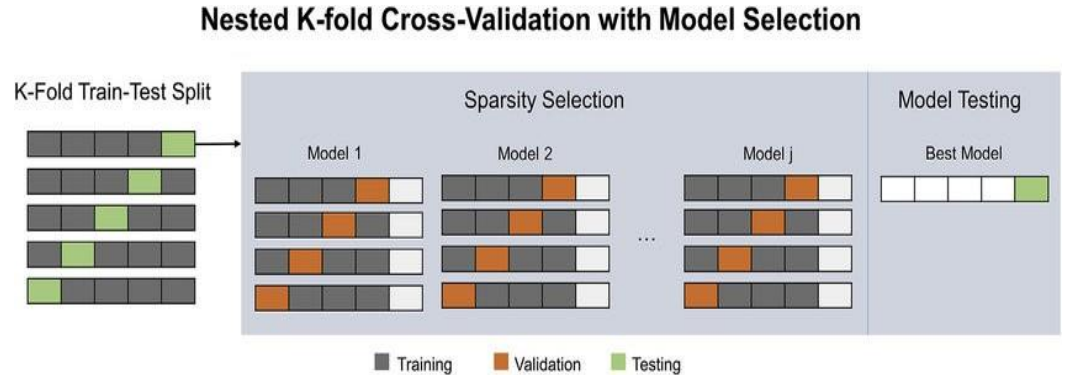

Fig 4:- Nested K-fold Cross-Validation with Model Selection

- *Train/test split:* Split the dataset into two pieces, so that the model can be trained and tested on different data. Better estimate of out-of-sample performance, but still a ”high variance” estimate. It is useful due to its speed, simplicity, and flexibility.
- *K-fold cross-validation:* Systematically create ”K” train/test splits and average the results together. It provides even better estimate of out-of-sample performance but runs ”K” times slower than train/test split.

Decision tree algorithm pseudocode:

- Place the best attribute of the dataset at the root of the tree.
- Split the training set into subsets. Subsets should be made in such a way that each subset contains data with the same value for an attribute.
- Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.

Decision tree algorithm:

In decision trees[3], for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with record’s attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node.We continue comparing our record’s attribute values with other internal nodes of the tree until we reach a leaf node with predicted class value. As we know how the modeled decision tree can be used to predict the target class or the value.

D. Doctor Recommendation

The doctor recommendation system works similar to that of e-commerce recommender systems where goods are rec- ommended to users based on user ratings and reviews of that product given by previous users and other analysis factors. Based on that some mathematical analysis will be done at server side that majorly includes higher rating, lower price, good feedback and less value for maintenance and service.

Here, doctor recommendation module aims to predict doc- tors by taking into consideration the ratings and reviews provided by patients[5]. Patients are given a platform where they can provide ratings and comments to doctors based on their experience. These ratings are analysed and the combined review provided by different patients is used to predict the best doctors having the required specialization nearby. The prediction is based on several factors including distance, specialization area, reviews etc. An appropriate weightage is provided to all these factors inorder to make the most relevant prediction.

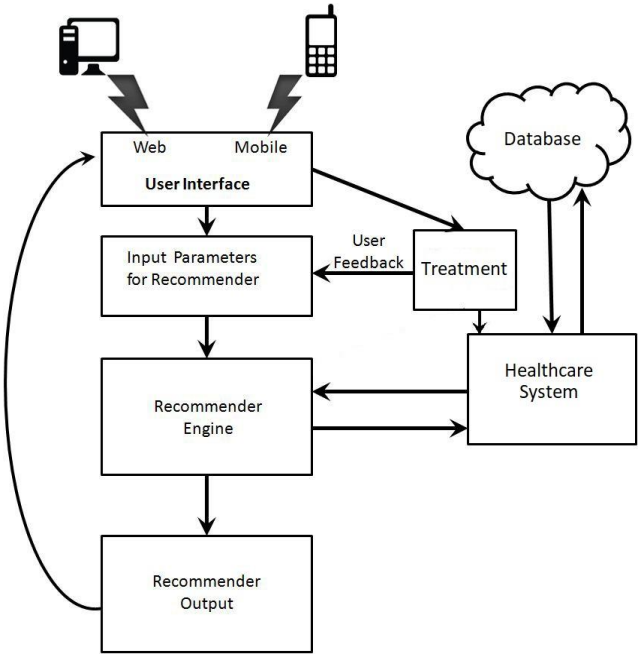

Fig 5:- Flowchart of doctor recommender system

ISSN No:-2456-2165

- *NLP to study patient reviews:* Patients provide a rating in the scale 1 to 5 to doctors. They can also write their reviews regarding each doctor, which are processed using TextBlob[6]. TextBlob is used to identify the sentiment and assign a polarity value in the range [-1,1] to the individual reviews provided by patients.

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. TextBlob is a python library that is built upon NLTK and Pattern. It offers a simple API to access its methods and perform basic NLP[11] tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. Sentiment analysis is basically the process of determining the attitude or the emotion of the writer, i.e., whether it is positive or negative or neutral. The sentiment function of textblob returns two properties, polarity, and subjectivity. Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float which lies in the range of [0,1].

The polarity of a word can be obtained using TextBlob(word).sentiment.polarity, which is a value between
-1 to 1 based on the sentiment of that word.

- *Category identification from symptoms:* Users are also given a provision to input their present conditions and symp- toms. Category to which that particular symptom belongs is identified. For example, if a user input is 'acute heart failure', then the machine predicts that the doctor in 'cardiology' specialization need to be consulted. And the prediction will be the best doctors of 'cardiology' department. Users can input multiple health conditions. In that case, the most department of doctor is identified by evaluating the severity of each condition. The most severe condition is identified first and the department corresponding to it will be prioritized.

- *Distance Calculation:* Another important factor that is considered while recommending doctor is the distance. If a user requests for a doctor, he is asked to input his location along with the specialization and other medical conditions. Doctor database hold location of all doctors. So using these locations, distance between the requested user and all doctors with that particular specialization is calculated as shown below.

Given two (latitude,longitude) coordinates, (lat1,lon1) and (lat2,lon2) in radians.

$$distance = 6371.01 * acos(sin(lat1) * sin(lat2) + cos(lat1) * cos(lat2) * cos(long1 - lon2)) \quad (1)$$

These distances help in predicting doctors with the best reviews and ratings nearby.

A weighted average rating value is calculated based on values of the above factors and names of few doctors with the highest weighted rating value is displayed, along with their details.

E. COVID-19 Chatbot

Coronavirus disease 2019 (COVID-19) is an infectious dis- ease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [7]. The disease was first identified in 2019 in Wuhan, the capital of Hubei China, and has since spread globally, resulting in the 2019–20 coronavirus pandemic.

The proposed healthcare system is embedded with a chatbot dedicated to COVID-19 risk assessment with a reliable patient screening protocol. It is developed using the Infermedica API. Using this chatbot an individual can undergo an interview by answering questions regarding his/her past medical conditions, present symptoms, travel history and finally after evaluating the individual's answers it provides various recommendations regarding the actions to take . Recommendations on what to do next are solely based on a combination of WHO(World Health Organisation) and CDC(The Centers for Disease Control and Prevention).

V. CONCLUSION

In the digital era, everything is available in our finger tips. But medical field is one area which has not undergone any such technological advances. That is the main concern behind the development of our proposed healthcare system. Entire medical history of a patient is made available at a single point, which is quite important for doctors in diagnosis.

Another major issue of the existing system is regarding the security of medical data stored in hospitals. This situation is overcome by making each person the only owner of their medical data, and they can decide who all can be given access to it. Disease prediction and doctor recommendation are the most exciting features, that is lacked in the existing system. Thus, the proposed healthcare system proves to be something that is quite relevant in today's world.

REFERENCES

[1]. Collier R. National Physician Survey: EMR use at 75% CMAJ. 2014;187(1):E17–8. Epub 2014 Dec 8.
[2]. T. Piliouras et al., ”Electronic health record systems: A current and future-oriented view,” 2013 IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, NY, 2013, pp. 1-6.
[3]. D. Dahiwade, G. Patle and E. Meshram, ”Designing Disease Prediction Model Using Machine Learning Approach,” 2019 3rd International Con- ference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 1211-1215.
[4]. S. Leoni Sharmila, C. Dharuman, P. Venkatesan, ”Disease Classification Using Machine Learning Algorithms - A Comparative Study”, Interna- tional Journal of Pure and Applied Mathematics, vol. 114, no. 6, pp. 1-10, 2017.
[5]. Yong-Feng Huang, Peng Liu, Qiao Pan and Jing-Sheng Lin, ”A doctor recommendation algorithm based on doctor performances and patient preferences,” 2012 International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP), Chengdu, 2012, pp. 92-95.
[6]. Textblob.readthedocs.io. 2020. Textblob: Simplified Text Process- ing — Textblob 0.15.2 Documentation. [online] Available at: https://textblob.readthedocs.io/en/dev/
[7]. COVID-19, MERS & SARS — NIH: National Institute of Allergy and Infectious Diseases, 2020
[8]. Fushiki, T. Estimation of prediction error by using K- fold cross-validation. Stat Comput 21, 137–146 (2011). https://doi.org/10.1007/s11222-009-9153-8
[9]. Gupta, Bhumika, Aditya Rawat, Akshay Jain, Arpit Arora and Naresh Dhami. “Analysis of Various Decision Tree Algorithms for Classification in Data Mining.” International Journal of Computer Applications 163 (2017): 15-19.
[10]. Seymour, Dr. Tom & Frantsvog, Dean & Graeber, Tod. (2014). Elec- tronic Health Records (EHR). 10.19030/ajhs.v3i3.7139.
[11]. Khurana, Diksha & Koli, Aditya & Khatter, Kiran & Singh, Sukhdev. (2017). Natural Language Processing: State of The Art, Current Trends and Challenges.