International Journal of Future Generation Communication and Networking
Vol. 13, No. 3s, (2020), pp. 110–120

Disease Risk Prediction Using Data Mining with Privacy Preservation of Data

Rutvik Mahajan, Parnal Tambat, Darshan Shinde, Kaustubh Yewale, Pooja Vengurlekar

Department of Computer Engineering, Sinhgad Institute of Technology and Science

rutvik.mj@gmail.com

tambatparnal2@gmail.com

darshanshinde235@gmail.com

kaustubhyewale88888@gmail.com

pnvengurlekar_sits@gmail.com

Abstract

Data mining-driven disease risk prediction has become one of the important topics in the field of e-healthcare. With the widespread use of hospital information system, there is a huge amount of generated data which can be used to improve healthcare service. However, without the security and privacy assurances, disease risk prediction cannot continue to flourish. To address this challenge, an efficient and privacy-preserving disease risk prediction model for e-healthcare is proposed. Compared with the up-to-date works, the proposed work comprehensively achieves two phases of disease risk prediction - disease model training and disease prediction, while ensuring privacy preservation. NAIVE BAYES algorithm is introduced to compute the classification result. The model makes use of the ABE algorithm for encryption and also demonstrates prediction of disease with the AID data set using the SVM algorithm. Besides, extensive performance evaluations demonstrate that our proposed model attains outstanding efficiency advantage and hence is more suitable for real-time e-healthcare, especially medical emergency.

Keywords— Disease prediction, Security, Data Mining, Healthcare service, Hospital management.

I. INTRODUCTION

Data mining is the core step, which results in the discovery of hidden but useful knowledge from massive databases. The discovered knowledge can be used by healthcare administrators to improve the quality of service. Health care data is massive. It includes patient-centric data, resource data, and transformed data. Health care organizations must have the ability to analyse data. Treatment records of millions of patients can be stored and computerized and data mining techniques may help in answering several important and critical questions related to health care. However, concerns are growing that the use of this technology can violate individual privacy. In recent years, wide available personal data has made privacy-preserving data mining an important issue. In the existing medical systems, there is a drawback of privacy related to patient's information. So, it is necessary to ensure patients feel fully confident to use the system and have their privacy control over it.

The remainder of this paper is organized as follows. In the rest of the section, we will discuss the motivation, the review of the literature, the proposed system, system architecture, future work and the conclusion.

ISSN: 2233-7857 IJFGCN

II. MOTIVATION

The motivation behind this is to handle a huge amount of different disease data and on that, the risk prediction of disease will be examined. With the widespread use of hospital information system, there is a huge amount of generated data which can be used to improve health care services. When it comes to medical and health care issues, the most important factor in preserving patient's sensitive data, thus developing data mining applications to provide people more customized health care services.

III. LITERATURE REVIEW

Literature survey is the most important step in any kind of research. Before starting developing, we need to study the previous papers of our domain which we are working and based on study we can predict or generate the drawback and start working with the reference of previous papers.

In this section, we briefly review the related work on An Efficient and Privacy-Preserving Disease Risk Prediction Scheme.

In this paper, the author proposes to give an idea about providing privacy to the historic medical data. It contains a privacy preserving patient decision support system which allows service provider to diagnose patient's disease without leaking any patient's historical medical data. The system model is divided into five parties: Trusted Authority (TA), Cloud Platform (CP), Data Provider (DP), Processing Unit (PU), and Undiagnosed Patient (PA). To prevent individual historic sensitive medical data to disclose from service provider. A new aggregation technique called additive homomorphic proxy aggregation (AHPA) scheme is introduced. To securely aggregate the message to solve the collusion problem, it contains the following six algorithms: KeyGen, ReKeygen, Encrypt, Decrypt, Re-encrypt & amp; Agg, and Re-decrypt. This algorithm (AHPA) can be applied in our Disease Risk Prediction Application to avoid the disclosure of patient's sensitive medical data without compromising the privacy of data provider. Since all the data is processed in the encrypted form our prediction application can achieve patient diagnose result in privacy preserving way [2].

In this paper, the author suggests the MHN architecture and privacy preserving data aggregation scheme. Qop can achieve authentication, guarantee integrity. the paper identifies the privacy requirements from the perspective of Qop. This paper describes the schema Encrypting the data prior to uploading it with symmetric encryption. This will be used to provide the privacy through symmetric encryption. In this technique the project uses various encryption techniques for data privacy. Few of them are: Encrypting the data prior to uploading it with some symmetric encryption; Using a Trusted Execution Environments (TEE) such as OS containers. Mainly the project is focusing on the first technique. It includes two type of encryption methods Symmetric and Asymmetric key encryption. In symmetric we make use of single key for both encryption and decryption process. Both sender and receiver have a copy of same key and the algorithm used is AES. In asymmetric we make use of two keys i.e. Public and Private keys. Public key is used for encryption and private used for decryption. Usually sender will have a public key and sender will have private key [3].

In this paper, the author suggests that by utilizing a set of similar random numbers generated from the human body properties such as the inter pulse interval (IPI) at different sites to encrypt and decrypt the symmetric key, the secure communication channel between body sensors is established. PHI names are semantically related to the PHI content itself, which breaks both the PHI confidentiality and the requester's interest privacy. The unauthorized persons can derive some privacy information with respect to the patient's interest. As a future scope, Sensors can be used to transfer precise medical information of patient to Medical service provider [4].

ISSN: 2233-7857 IJFGCN

In this paper, the author suggests an idea about the privacy preserving scheme for medical data. It makes use of SVM algorithm which is one of the powerful classification algorithms in terms of prediction. This paper describes a model similar to our application in which data privacy is done using encryption and decryption techniques. This paper aims at the accurate results, low computation and data privacy which is similar to our application. The encryption and decryption techniques can be applied in our Disease Risk Prediction Application to avoid the disclosure of patient's sensitive medical data also without compromising the privacy of data provider. Since the identity of the users is in encrypted form, privacy is preserved and also the user can get efficient medical pre-diagnosis results.[5].

In this paper, the author has presented an intelligent and effective heart attack prediction methods using data mining. Firstly, it provided an efficient approach for the extraction of significant patterns from the heart disease data warehouses for the efficient prediction of heart attack Based on the calculated significant weight age, the frequent patterns having value greater than a predefined threshold were chosen for the valuable prediction of heart attack. In this paper the drawbacks are for predicting heart attack significantly 15 attributes are listed. Besides the 15 listed in medical literature we can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules. [6]

In this paper, author design an inference attack-resistant e-healthcare cloud system with fine-grained access control. We first propose a two-layer encryption scheme. To ensure an efficient and fine-grained access control over the EHR data, we design the first-layer encryption, where we devise a specialized access policy for each data attribute in the EHR, and encrypt them individually with high efficiency. To preserve the access pattern of data attributes in the EHR, we further construct a blind data retrieving protocol. We also demonstrate that our scheme can be easily extended to support search functionality. Finally, we conduct extensive security analyses and performance evaluations, which confirm the efficacy and efficiency of our schemes. [7]

In this paper the author proposes a semantic-based secure discovery framework for mobile healthcare enterprise networks that exploits semantic metadata (profiles and policies) to allow flexible and secure service search/retrieval. As a key feature, this approach integrates access control functionalities within the discovery framework to provide users with filtered views on available services based on service access requirements and user security credentials. Identification of solutions to these challenges is critical if clinical decision support is to achieve its potential and improve the quality, safety and efficiency of healthcare. [8]

In this paper the author proposed a method that, given a query submitted to a search engine, suggests a list of related queries. The related queries are based in previously issued queries, and can be issued by the user to the search engine to tune or redirect the search process. The method proposed is based on a query clustering process in which groups of semantically similar queries are identified. The clustering process uses the content of historical preferences of user's registered in the query log of the search engine. The method not only discovers the related queries, but also ranks them according to a relevance criterion. Finally, we show with experiments over the query log of a search engine the effectiveness of the method. [9]

In this paper, the author proposed Lightweight Sharable and Traceable, a lightweight secure data sharing solution with traceability for mHealth systems. Lightweight Sharable and Traceable seamlessly integrates a number of key security functionalities, such as fine-grained access control of

ISSN: 2233-7857 IJFGCN

encrypted data, keyword search over encrypted data, traitor tracing, and user revocation into a coherent system design. Considering that mobile devices in mHealth are resource constrained, operations in data owners' and data users' devices in Lightweight Sharable and Traceable are kept at lightweight and provide security. Further, extensive experiments on its performance (on both PC and mobile device) demonstrated that Lightweight Sharable and Traceable is very promising for practical applications. [10]

IV. GAP ANALYSIS

Compared with the existing paper in this paper, an efficient and privacy-preserving disease risk prediction scheme for e-healthcare is proposed. In the existing paper there is a drawback of security related to patient's information. So, in the proposed work we are going to use encryption technique to provide security to the sensitive information of the patients. Compared with the existing work we are going to use Naïve Bayes algorithm to search the data, and used encryption algorithm to provide security and SVM algorithm for predict the diseases.

V. PROPOSED SYSTEM

The proposed system we build which leverages data mining methods to reveal the relationship between the regular physical examination records and the health risk given by the user. Data Mining algorithms like Naïve Bayes, Support Vector Machine are used for the disease prediction and for the storage of the data the system used the MYSQL database. The system provides a user-friendly interface for various users and doctors. In this paper, an efficient and privacy-preserving disease risk prediction scheme for e-healthcare is proposed. In the existing paper, there is a drawback of security related to the patient's information. So, in the proposed work we are going to use encryption technology to provide security to the sensitive information of the patients. Compared with the existing work we are going to use the Naïve Bayes algorithm to search the data and used encryption algorithm to provide privacy.

The proposed system comprises of following stages: (i) Login to the System, (ii) Search Doctor/Hospital, (iii) Enter Symptoms, (iv) Processing, and (vi) Predict Disease.

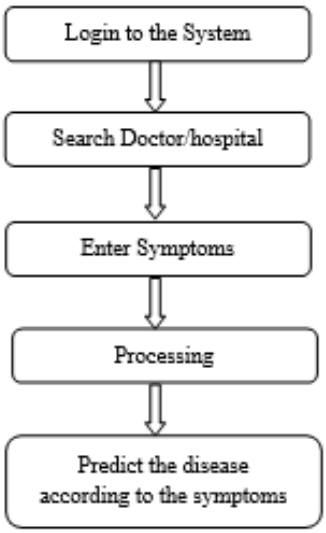

Fig. 1 Flow Chart

ISSN: 2233-7857 IJFGCN

VI. SYSTEM ARCHITECTURE

The system comprises of 3 parts. The pre-processing part, the logical part, and the user part. The pre-processing part has options to search for the name of the hospital or doctor and store the result in the database as per the result. The symptoms given by the user are then given to the logical part. In the logical section, the disease is predicted with the help of machine learning algorithms. The user part has the following functions- he/she can search for any hospital nearby, he/she can get the nearby hospital as per the current location.

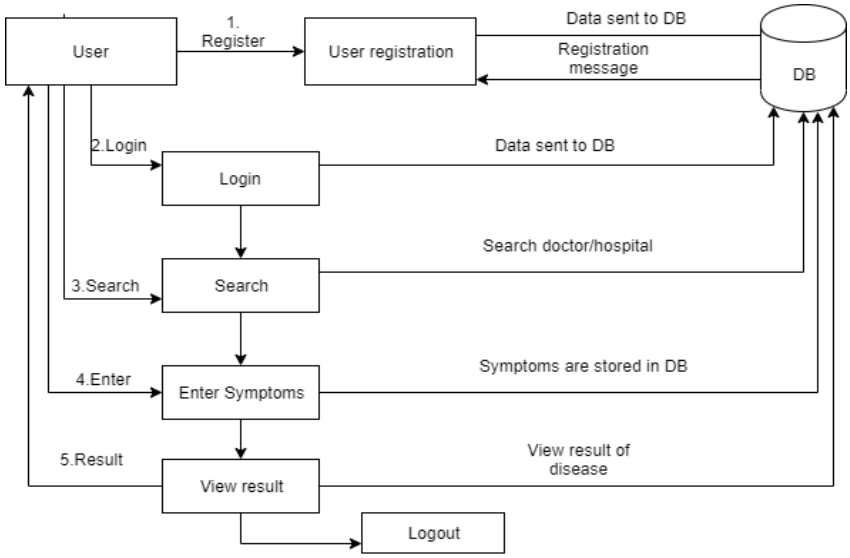

Fig. 2 System Architecture (User)

User can register if he/she is a new user. If already registered, the user can directly log in. The data is stored in DB. After login, the user can search the doctor or hospital. Next is to enter symptoms according to the user's conditions. The result is displayed at the last to the user.

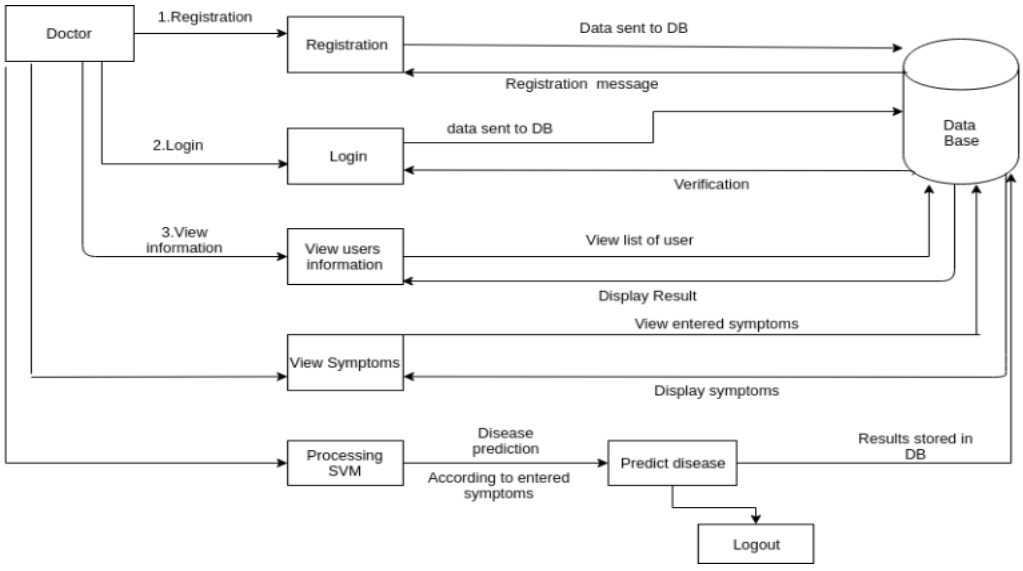

Fig. 3 System Architecture (Doctor)

ISSN: 2233-7857 IJFGCN

Doctor can register to the system. The doctor can log in and the verification message is sent to the doctor in return. The doctor can view the user's symptoms which are entered by the users. By using the SVM algorithm for prediction we can predict the disease in the module.

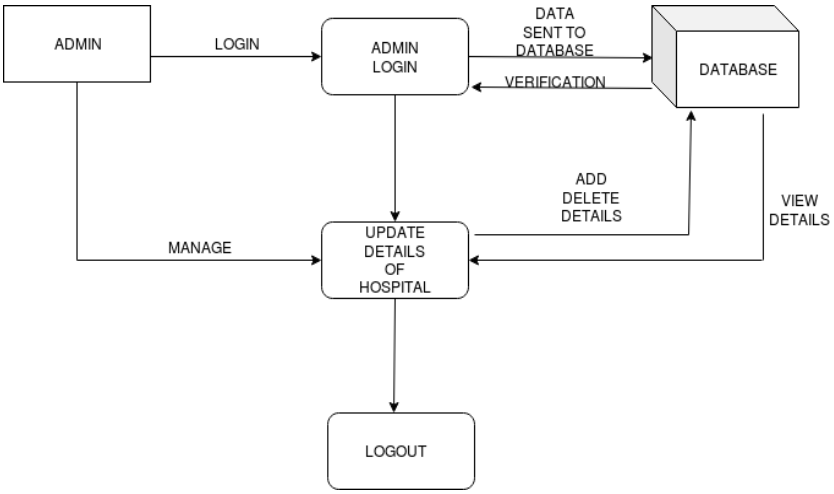

Fig. 4 System Architecture (Admin)

Admin is used to manage the database and the other two modules (Doctor & User). Admin can update hospital details, or delete certain details and also can view them. Admin can log in and logout.

VII. Result and Discussion

Experimental results are discussed below

A. *User Interface*:

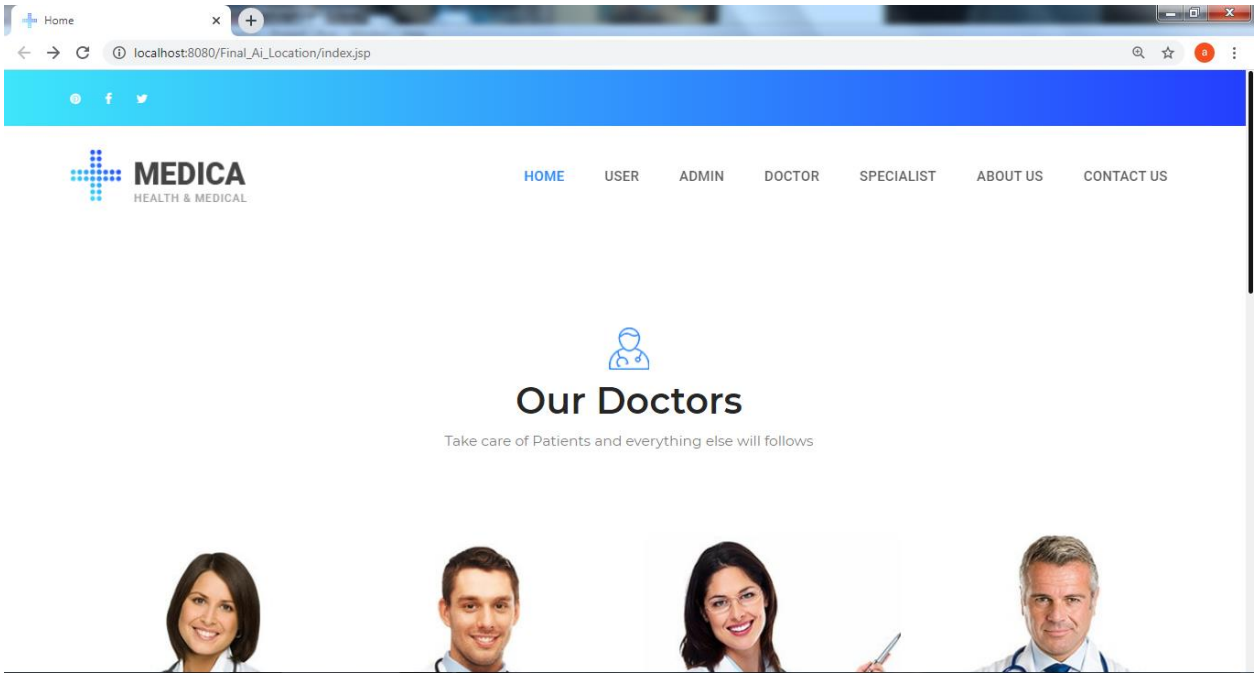

Fig. 4: User Interface showing details of Homepage

ISSN: 2233-7857 IJFGCN

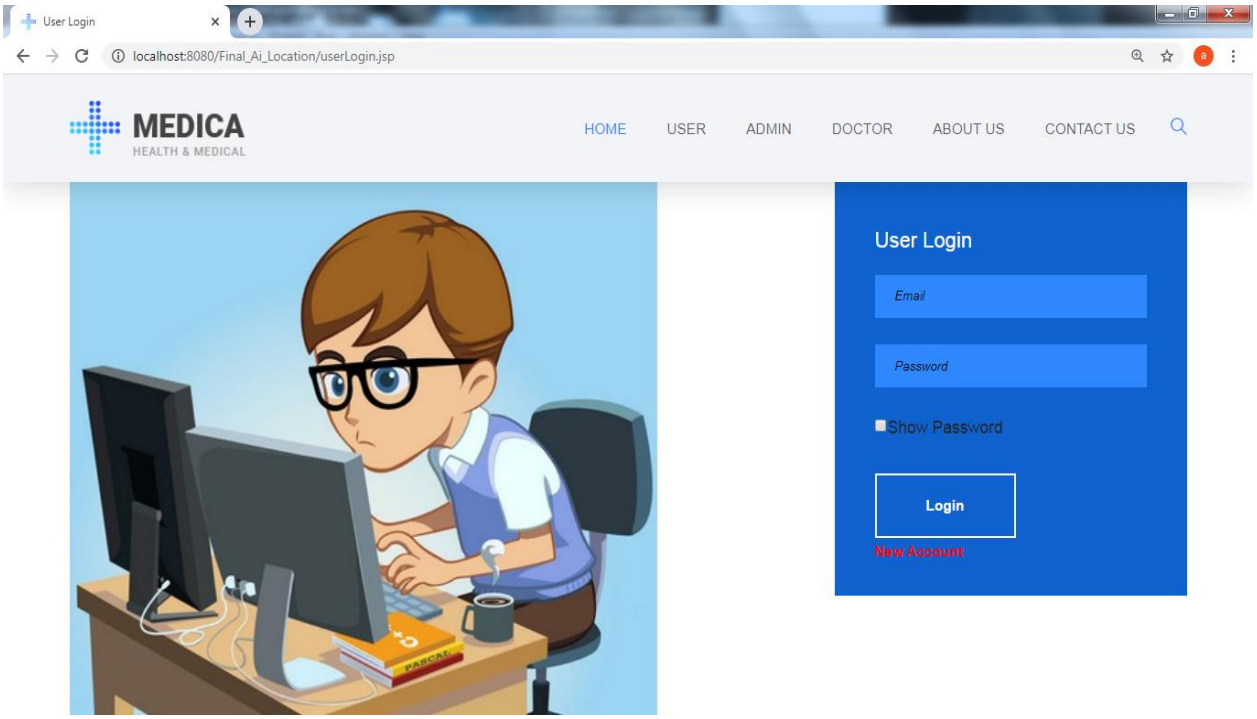

Fig. 5: User Interface showing details of user, doctor and admin login

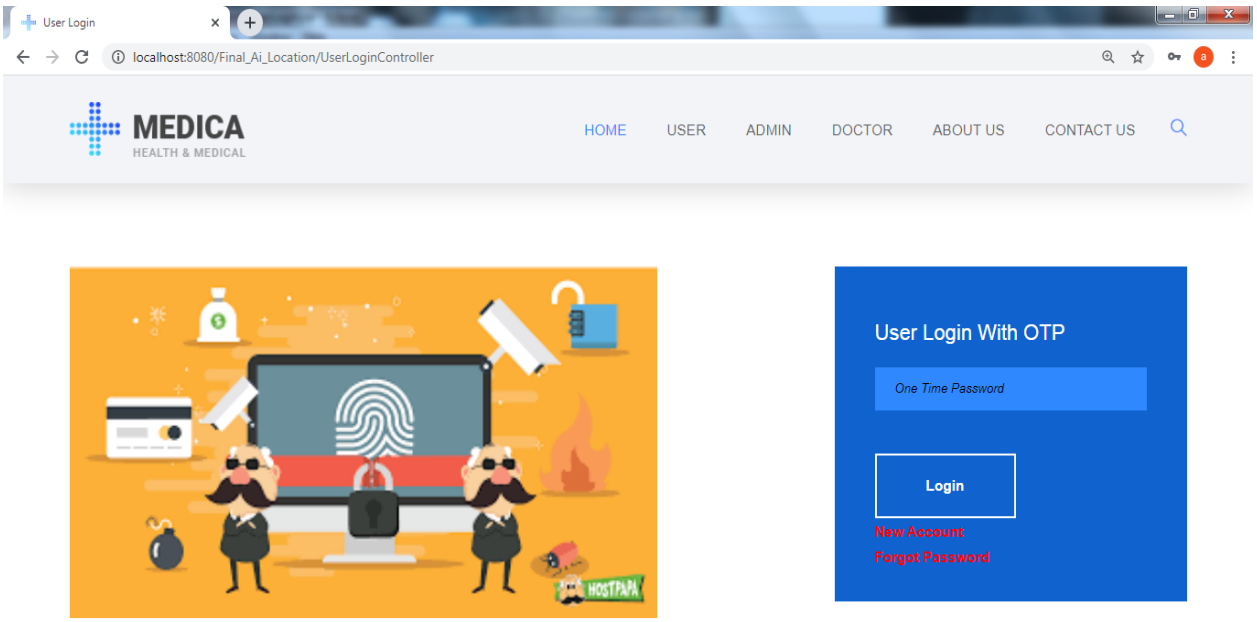

Fig. 6: User Interface showing details of user login with OTP

ISSN: 2233-7857 IJFGCN

B. Disease Prediction:

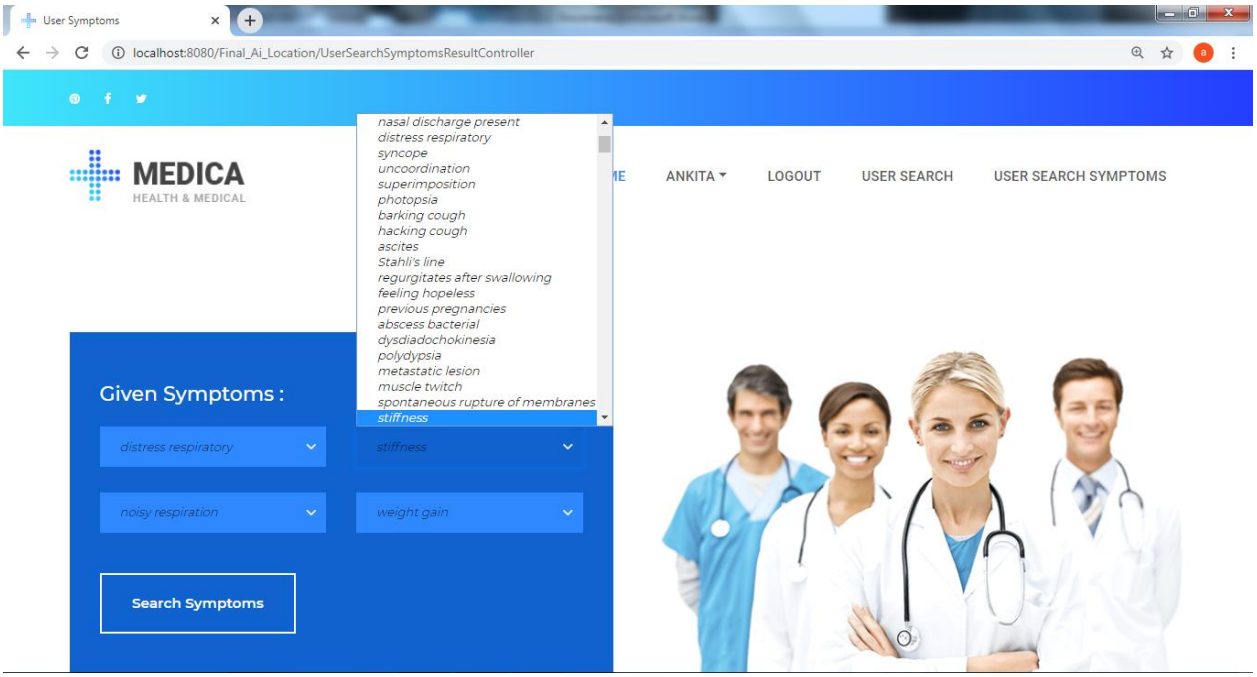

Fig. 7: Search symptoms

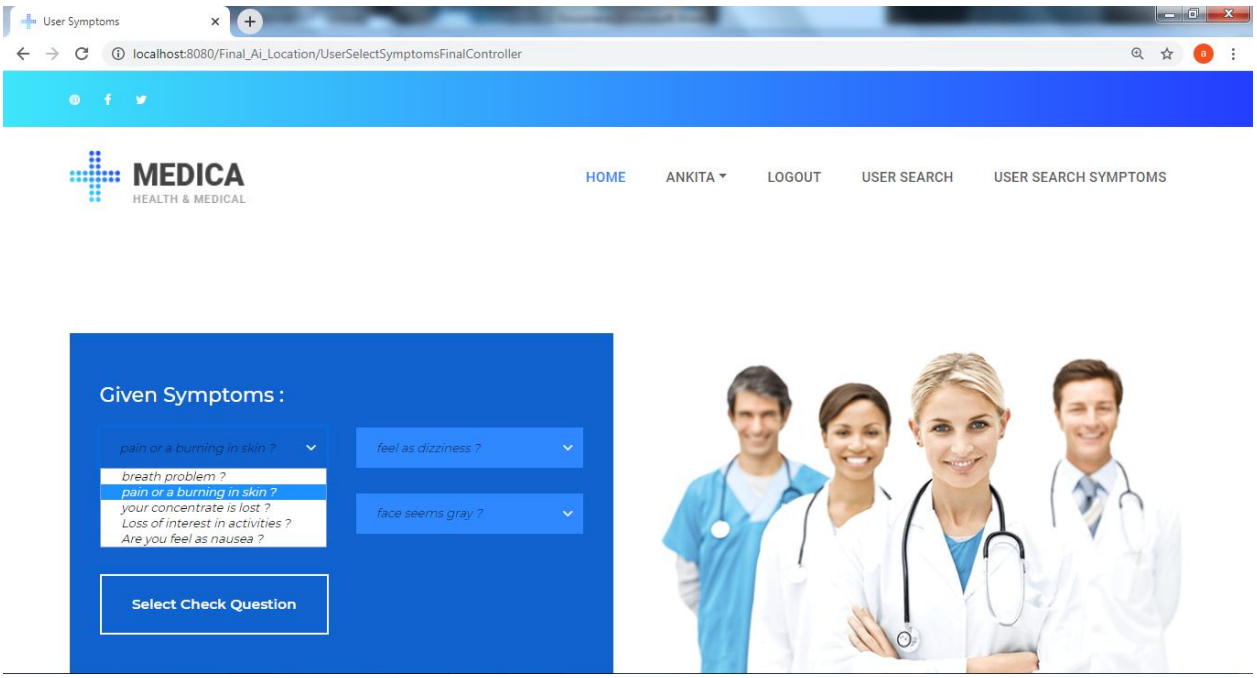

Fig. 8: Predicting disease according to symptoms

ISSN: 2233-7857 IJFGCN

C. Backend Details:

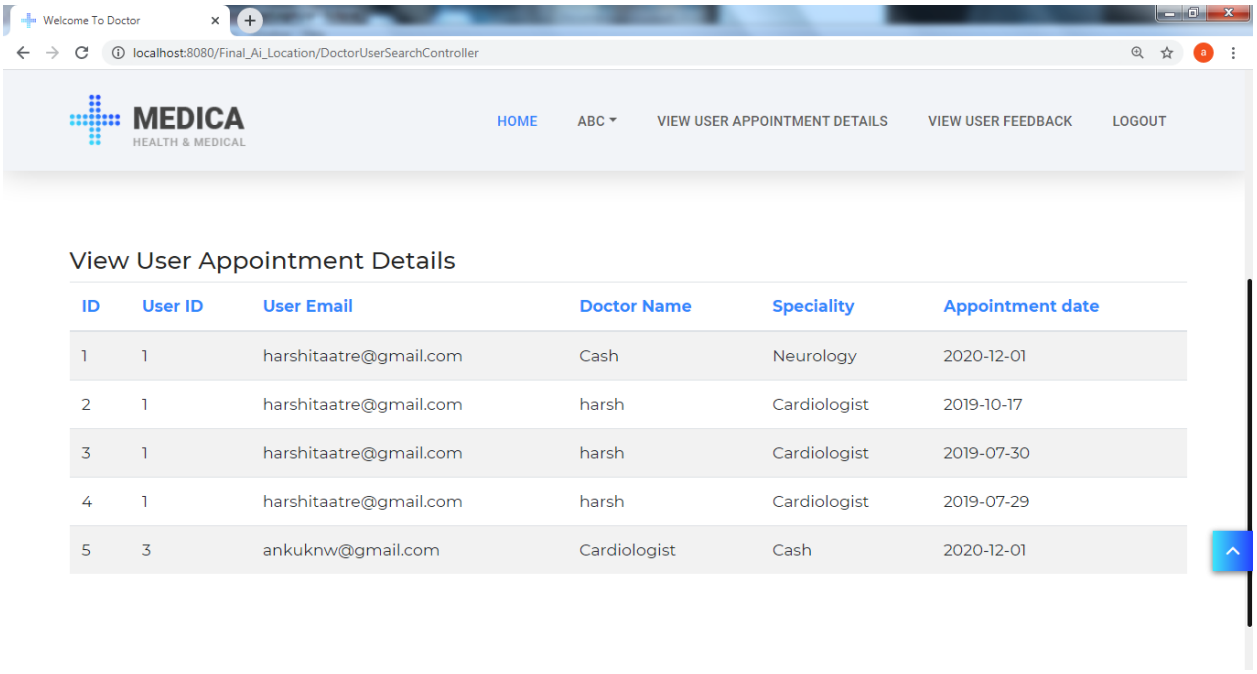

Fig. 9: User appointment details

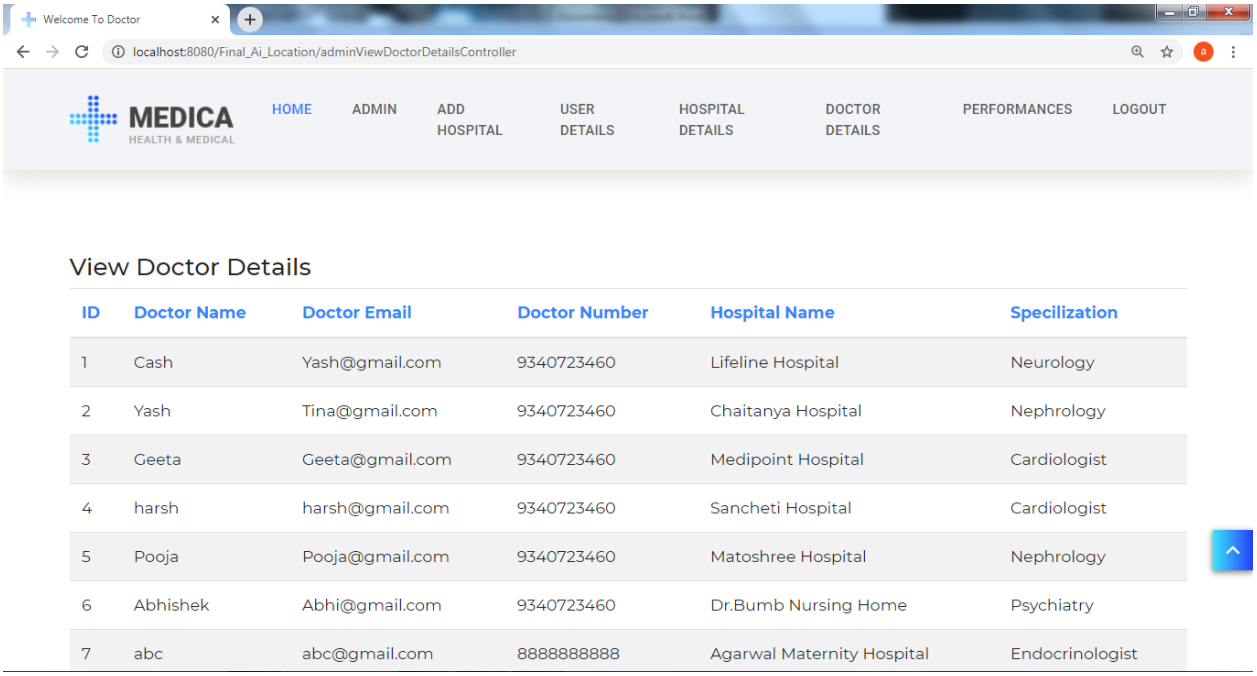

Fig. 10: Doctor details

ISSN: 2233-7857 IJFGCN

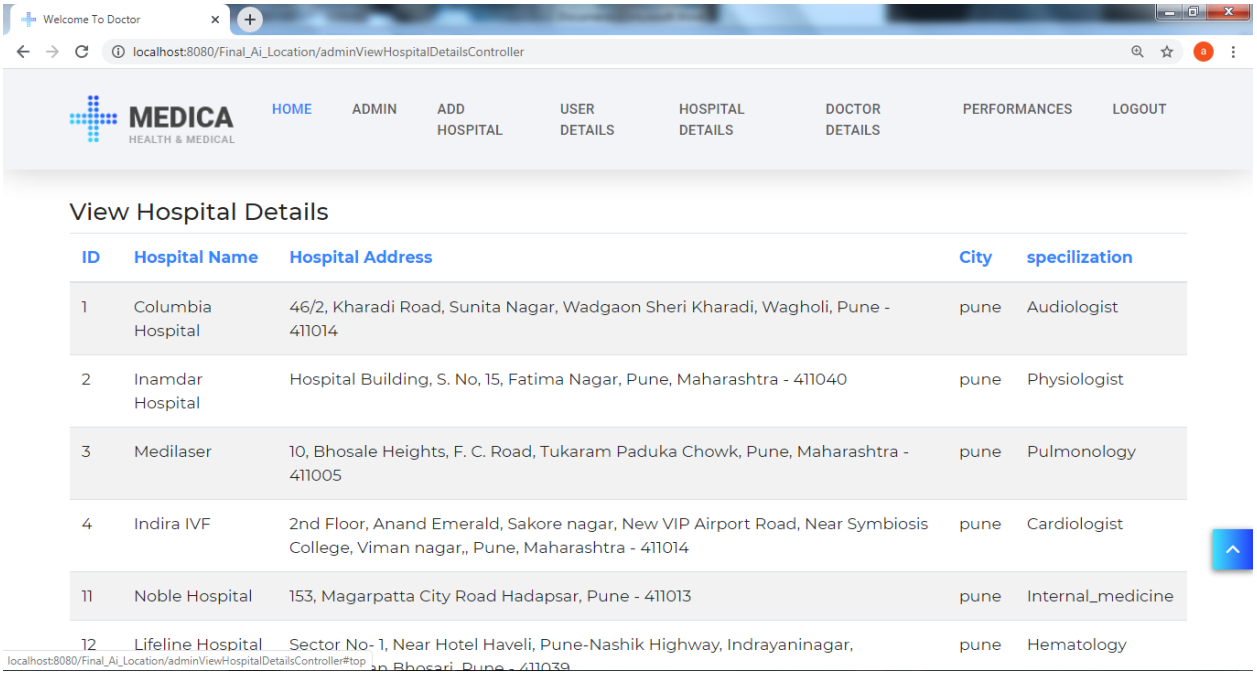

Fig. 11: Hospital details

VIII. FUTURE SCOPE

This paper focuses on prediction of the disease. In near future, application can be advanced by, achieving the message integrity in e-healthcare since it directly influences the accuracy of diagnosis results and even the life safety of patients.

IX. CONCLUSION

Physicians, most healthcare services do not provide us with accurate results. The project develops a system that can predict and examine the diseases based on the symptoms and medical history. In the system, the user or patient can search the hospital or doctor. Users/ Patients give the symptoms and the system predicts diseases and medicines. The project focuses on preserving patients' sensitive data which is essential in E-Health Care systems as it directly influences the life safety of patients.

REFERENCES

[1] Xue Yang, Rongxing L,``An Efficient and Privacy-Preserving Disease Risk Prediction Scheme for E-Healthcare",2018.
[2] Ximeng Liu, Rongxing Lu, Jianfeng Ma, “Privacy Preserving Patient Centric Clinical Decision Support System on Naive Bayesian Classification”,2015.
[3] Kuan Zhang, Kan Yang, Xiaohui Liang, Zhou Su, “Security and Privacy for Mobile Healthcare Networks: From a Quality of Protection Perspective”,2015.
[4] Hui Zhu,Xiaoxia Liu, “Efficient and Privacy-Preserving Online Medical Pre-Diagnosis Framework Using Nonlinear SVM”,2015.
[5] Yonglin Ren, Richard Werner Nelem Pazzi, And Azzedine Boukerche, “Monitoring Patients Via a Secure and Mobile Healthcare System”,2010.
[6] Srinivas K, Rani B K, Govrdhan A. “Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks”. International Journal on Computer Science & Engineering, 2010.
[7] Ndibanje Bruce, Mangal Sain, Hoon Jae Lee, “A Support Middleware Solution for e-Healthcare System Security”, IEEE 16th International Conference on Advanced Communication Technology.

ISSN: 2233-7857 IJFGCN

[8] Wei Zhang, Yaping Lin, Jie Wu, Fellow and Ting Zhou "Inference Attack-Resistant E-Healthcare Cloud System with Fine-Grained Access Control", IEEE Transactions on Services Computing 2018.
[9] R. Baeza-Yates, C. Hurtado, and M. Mendoza, "Query recommendation using query logs in search engines," in Proc. Int. Conf. Current Trends Database Technol., 2004, pp. 588–596.
[10] Yang Yang, Ximeng Liu, Robert H. Deng, Yingjiu Li, "Lightweight Sharable and Traceable Secure Mobile Health System", IEEE Transactions on Dependable and Secure Computing, 2017.

ISSN: 2233-7857 IJFGCN