# An Efficient and Privacy-Preserving Disease Risk Prediction Scheme for E-Healthcare

VARSHA PRASHANT JADHAV

PG Student, Department of Information Technology, Bharati Vidyapeeth (Deemed to be university) College of Engineering Pune.

PROF. PRAKASH .R. DEVALE , PROF. PRAMOD A. JADHAV

Department of Information Technology, Bharati Vidyapeeth (Deemed to be university) College of Engineering Pune.

## Abstract

Data analysis plays a significant role in handling a large amount of data in the healthcare. The previous medical researches based on handling and assimilate a huge amount of hospital data instead of prediction. Due to an enormous amount of data growth in the biomedical and healthcare field the accurate analysis of medical data becomes propitious for earlier detection of disease and patient care. With the development of society and economy, people pay more attention to their own health. The demand of more personalized health service is gradually rising. However, due to the lack of experienced doctors and physicians, most healthcare organizations cannot meet the medical demand of public. Due to that public want the Medical Treatment online with accuracy. Now a day's, public has no time to get the doctor physically, and then search the online hospital near about the current location. With the widespread use of hospital information system, there is huge amount of generated data which can be used to improve healthcare service. Thus, more and more data mining applications are developed to provide people more customized healthcare service. In the proposed paper we use different algorithm to increase the security of sensitive information of hospital management includes doctors, patients and so on.

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

## Introduction

Mostly 50% of peoples suffer from one or more than one chronic disease and because of that, those spent more amounts on the treatment of disease. As the lifestyle is improved, then the frequency of disease is also increasing. In India, nearly 61% of death causes due to the non – communicable disease like heart disorder, cancer, and diabetes. The main reason behind the cause of diseases is environmental condition and living habits of people. To get earlier disease detection, reduced the risk of disease and diagnosis of the disease there is IOT based disease prediction used Knowledge discovery in databases is well-defined process consisting of several distinct steps. Data mining is the core step, which results in the discovery of hidden but useful knowledge from massive databases. Data mining technology provides a user-oriented approach to novel and hidden patterns in the data. The discovered knowledge can be used by the healthcare administrators to improve the quality of service. The discovered knowledge can also be used by the medical practitioners to reduce the number of adverse drug effect, to suggest less expensive therapeutically equivalent alternatives. Anticipating patient's future behavior on the given history is one of the important applications of data mining techniques that can be used in health care management. A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Health care data is massive. It includes patient centric data, resource management data and transformed data. Health care organizations must have ability to analyze 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. In this context, electronic healthcare systems (EHRs) employee such rules and thus were categorized as security critical systems. These systems are differentiated in one important aspect to other systems, the balancing between confidentiality and availability. The tension between these goals is clear: while all the patient's data should be available to be shared and monitored to deliver professional healthcare services; for security reasons, part of the data may be considered confidential and must not be accessible. In EHRs, users may be a health data owner (i.e., patients) or a requester (i.e., doctors or pharmacists), servers, in turn could be local or cloud servers that store, process and analyze the gathered health data. Networks, on the other hand, act as the bridge connecting between patients and the medical staff to support the transmitting and sharing of data. So, it is necessary to ensure patients feel fully confident to use the system and have their own privacy control over it. To this end, in this paper, we conduct an in-depth survey study to analyze the healthcare system's security and privacy threats. Then, we propose a novel security model that captures the scenario of data interoperability and supports the security fundamental of EHR along with the capability of providing fine-grained access control.

The reminder of this paper is organized as follows. In rest of the section we will discuss about the motivation, the review of literature, problem statement, design proposed system, and finally, concludes the paper.

**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, thus developing data mining applications to provide people more customized health care service.

**Related Work**

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of 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 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. [1]

In this paper, author presented a middleware solution approach to support data and network security over e-Healthcare system sing medical sensor networks. It has been shown that a masquerade attack can be launched to the system and patients 'data are in danger. We proposed this middleware to counter this kind of attack where a user and all devices into the healthcare network are mutual authenticated. Finally a performance analysis has been done with regard to masquerade attack and the result reveals the efficient of the proposed solution. [2]

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. [3]

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. [4]

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. [5]

The various heart disease prediction techniques are discussed and analyzed in this paper. The data mining techniques used to predict heart diseases are discussed here. Heart disease is a mortal disease by its nature. This disease makes several problems such as heart attack and death. In the medical domain, the significance of data mining is perceived. From the comparative study we can conclude that Support Vector Machine (SVM) technique is an efficient method for predicting heart disease. [6]

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 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. [7]

In this paper, the author proposes a Highly Available, Scalable and Secure distributed data storage system for high performance and secure data management. Distributed and parallel data storage or file systems such as Object-based Storage Devices and flexible key distribution schemes Data at rest (static) and in transit (dynamic) are protected with different encryption strategies for privacy and integrity. Secret sharing and replication support both security and availability. Encryption and key management are not necessary in data at rest protection. The future work includes a detailed simulation and further performance analysis. [8]

In this the author proposed a security scheme for users. This scheme provides storing and sharing their intricate data in the Cloud environment. This scheme provides vital encryption and decryption technique for achieving security on cloud application. The revocation procedure is an explicit performance destroyer within the access control method in cryptography. In this scheme, the unique data is firstly separated into numerous parts. Then these parts are sent to the cloud server. Whenever a user revocation happens, the data owner desires merely to retrieve one part and re-encrypt it. This scheme is based on cryptographic storage application. Furthermore techniques are implemented to improve the security of the data. [9]

In this paper, the author have proposed a protected multiple owner data sharing system. This system is used for dynamic groups in the cloud environment. Any Cloud user can unidentified person distribute data with other users in order to improve the signature of group and dynamic broadcast encryption techniques. For this, the storage transparency and encryption calculation cost are self governing with respect to the number of users that are revoked. Furthermore, the protection and investigation system with exact proofs is analyzed. [10]

### Problem Statement

A feedback mechanism could save manpower and improve performance of system automatically. The doctor could fix prediction result through an interface, which will collect doctors' input as new training data. An extra training process will be triggered everyday using these data. Thus, our system could improve the performance of prediction model automatically.

### Proposed Method

The proposed system we build which leverages data mining methods to reveal the relationship between the regular physical examination records and the potential health risk given by the user or public. This system uses the Machine learning and 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 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 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.

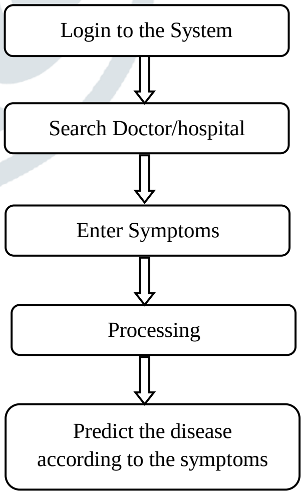

**Fig.1 Flow diagram**

## Architecture

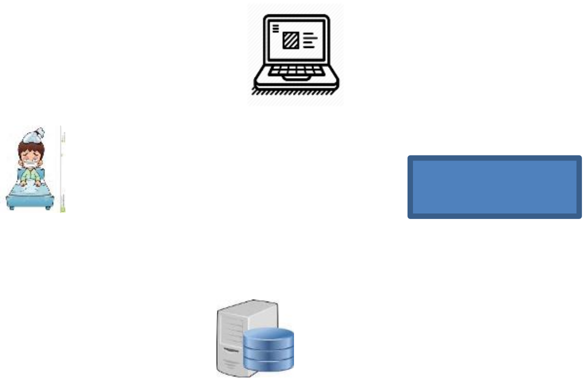

Fig.2 System Architecture

## Conclusion

In this paper, we proposed a privacy-preserving disease predicting system which can help physicians make a proper diagnosis of disease and provide health services for patients anytime anywhere in a privacy-preserving way. This project implements a disease risk prediction system which leverages data mining methods to reveal the relationship between the regular physical examination records and the potential health risk given by the user or public. Different machine learning algorithms are applied to predict physical status of examinee that will be in danger of physical deterioration next year. In our approach user or patient search the hospital and the results given are according to the nearest location of current location of user/patient. User / Patient give symptoms and the system predicts the disease and provides the medicines. The benefits of privacy-preserving diagnosis are to maintain the balance between security and efficiency which should be considered firstly. Therefore, how to optimize the model training using set for efficiency improvement and finding an effective way of introducing some other advanced machine learning methods to build the privacy-preserving disease prediction system are worthy of investigation.

## References

[1] 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.

[2] Alessandra Toninelli, Rebecca Montanari, And Antonio Corradi “Enabling secure service discovery in mobile healthcare enterprise networks”, IEEE Wireless Communications Volume: 16, Issue: 3 , June 2009.

[3] Ndibanje Bruce, Mangal Sain, Hoon Jae Lee, “A Support Middleware Solution for e-Healthcare System Security”, IEEE 16th International Conference on Advanced Communication Technology.

[4] 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.

[5] 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.

[6] Cincy Raju, Philipsy E, Siji Chacko, L Padma Suresh, Deepa Rajan S, “A Survey on Predicting Heart Disease using Data Mining Techniques”, 2018 Conference on Emerging Devices and Smart Systems (ICEDSS).

[7] 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.

[8] Zhiqian Xu, Hai Jiang, “HASS: Highly Available, Scalable and Secure Distributed Data Storage Systems”, 2009 International Conference on Computational Science and Engineering.

[9] Kamara, S., Lauter, K. Sion, R., Curtmola, R., Dietrich, “Cryptographic Cloud Storage”, 2010 Workshops of LNCS Springer, Heidelberg, vol. 6054, pp. 136-149, 2010.

[10] Bethencourt, J., Sahai, A.,Waters, B., “Ciphertext policy attribute-based encryption”, 28th IEEE Symposium on Security and Privacy, pp. 321-334, 2007.

[11] J. Zhou, X. Lin, X. Dong, and Z. Cao, “PSMPA: patient self controllable and multi-level privacy-preserving cooperative authentication in distributed m-healthcare cloud computing system,” IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 6, pp. 1693–1703, 2015.

[12] G. Wang, R. Lu, and C. Huang, “PSLP: privacy-preserving single-layer perceptron learning for e-healthcare,” in 10th International Conference on Information, Communications and Signal Processing, ICICS 2015, Singapore, December 2-4, 2015, 2015, pp. 1–5.

[13] J. Vaidya, M. Kantarcioglu, and C. Clifton, “Privacy-preserving naive bayes classification,” VLDB J., vol. 17, no. 4, pp. 879–898, 2008.

[14] T. Shemeikka, P. Bastholm-Rahmner, C. Elinder, A. Veg, E. Tornqvist, B. Cornelius, and S. Korkmaz, “A health record integrated clinical decision support system to support prescriptions of pharmaceutical drugs in patients with reduced renal function: Design, development and proof of concept,” I. J. Medical Informatics, vol. 84, no. 6, pp. 387 395, 2015.

[15] D. Boneh and M. K. Franklin, “Identity-based encryption from the weil pairing,” in Advances in Cryptology - CRYPTO 2001, 21st Annual International Cryptology Conference, Santa Barbara, California, USA, August 19-23, 2001, Proceedings, 2001, pp. 213–229.