

Clinical Decision Support System for Patient Centric System in Privacy Preserving Way Using Nave Bayesian Classification

1. Manodnya A. Shitole., *Amrutvahini College of Engineering, Sangamner*
2. Prof. Manoj A. Wakchaure, *Amrutvahini College of Engineering, Sangamner*

Abstract—A clinical decision support system forms critical capability to link health observations with health knowledge to influence choices by clinicians for improved healthcare. Clinical decision support system, having data mining technique which helps us for extracting data which we want. The clinical decision support system gives us the advantage which of provides the better diagnosis accuracy and also minimize the diagnosis time. The large amounts of clinical data generated every day by many of healthcare system so the naive Bayesian classification can be utilized to formed valuable information to improve clinical decision support system. But the clinical decision support system is quite promising but it also having many challenges about security of patient data. So for that I propose a new privacy-preserving patient-centric clinical decision support system, which helps clinician complementary to diagnose the risk of patients disease in a privacy-preserving way. As the past patients historical data are stored in cloud so patient can use it anywhere anytime and used the train naive Bayesian classifier for finding the top-k diseases without leaking any individual patient medical data. Specifically, to protect the privacy of past patients historical data, a new cryptographic tool called additive homomorphism proxy aggregation scheme is designed. Moreover, to leverage the leakage of nave Bayesian classifier, we introduce a privacy-preserving top-k disease names retrieval protocol in our system with document uploading facility. Detailed privacy analysis ensures that patients information is private and will not be leaked out during the disease diagnosis phase. In addition, performance evaluation via extensive simulations also demonstrates that our system can efficiently calculate patients disease risk with high accuracy in a privacy-preserving way.

Index Terms—Privacy Preserving, Clinical decision support system, naive Bayesian.

1 INTRODUCTION

IN Healthcare Industry, no appropriate technique is developed to find great potential economic values from big healthcare data, these data might not only become meaningless but also requires a large amount of space to store and manage. From last two decades the miraculous evolution in data mining technique has imposed a major impact on the revolution of humans lifestyle by predicting behaviors and future trends on everything which can convert stored data into meaningful information. To speed up the diagnosis time and improve the diagnosis accuracy, a new system in healthcare industry should be workable to provide a much cheaper and faster way for

diagnosis.[1] Clinical Decision Support System (CDSS), with various data mining techniques being applied to assist physicians in diagnosing patient diseases with similar symptoms, has received a great attention recently. Naive Bayesian classifier, one of the popular machine learning tools, has been widely used recently to predict various diseases in CDSS. Despite its simplicity, it is more appropriate for medical diagnosis in healthcare than some sophisticated techniques.

2 LITERATURE SURVEY

Data mining techniques have been widely used in clinical decision support systems for prediction and diagnosis of various diseases with better accuracy. The techniques have been very effective in developing clinical support

systems because they are able to detect hidden patterns and relationships in medical data. So here we review the different data mining techniques for clinical decision support system.

2.1 Bayesian Belief Network

The Bayesian network is knowledge based graphical representation that shows a set of variables and their probabilistic relationships between diseases and symptoms. Bayesian network is utilized to find the probability of the presence of possible diseases given their symptoms. Its advantage is that it requires the knowledge and conclusions of experts in the form of probabilities. It is very important for the physician who has no computer expertise to understand about the Bayesian network. Which is also gives as a clinician reference with a searchable database of diseases and clinical manifestations [5]. The another way Simul Consult is applies a statistical pattern-matching approach that considers the age of onset and offset of the findings in each disease.[6]

2.2 Decision Tree

Decision tree is the most often used techniques of data analysis. It is applied to classify records to a proper class. In medical field decision trees determine the sequence of attributes. Firstly it gives a set of solved cases. Then the whole set is divided into training set and testing set. A training set is used for the induction of a decision tree. A testing set is used to find the accuracy of an obtained solution. AY Al-Hyari developed a CDSS for diagnosing patients with Chronic Renal Failure using various classification methods like neural network, naive bays and decision tree. They proved that there is (92.2

2.3 Naive Bayes

Naive Bays uses the kernel estimator for numeric attributes rather than a normal distribution and utilized Supervised Discretization while converting numeric attributes to normal

ones. We got an Output in text form of Naive Bayes classifier. Abeer Y. Al-Hyari designed a CDSS for prediction and diagnosis of Chronic Renal Failure (CRF) using naive bays. The implemented CDSS can be used to observe the progression stage of the disease in patient. They were able to achieve accuracy of 88.2

3 EXISTING SYSTEM

Although clinical decision support system is quite promising and tremendously flourishing but still it faces many challenges including information security and privacy concerns and the files will taken with it [1]. Large amount of data store on server but this data are not store in encrypt format.

4 PROBLEM STATEMENT

The problem is to the determine how to securely store file on server and upload the documents for future use. Also, to develop a new privacy preserving patient centric clinical decision system, which help clinical complementary to diagnose the risk of patients diseases in a privacy preserving way with finding the top-k diseases names for patients use.

5 PROPOSED SYSTEM

To propose a new privacy-preserving patient-centric clinical decision support system, this helps clinician complementary to diagnose the risk of patients disease in a privacy-preserving way. In the proposed system the past data are stored in cloud and it can be used for naive Bayesian classifier to compute the disease risk for new coming patients and also patient can retrieve the top-k disease names according to their own preferences without leaking any individual patient medical data[7]. Detailed privacy analysis ensures that patients information is private and will not be leaked out during the disease diagnosis phase. In addition, performance evaluation via extensive simulations also demonstrates that our system can efficiently calculate patients disease risk

with high accuracy in a privacy-preserving way. Our key technical contribution is a privacy-preserving method that allows a data miner to compute frequencies of values or tuples of values in the customers' data, without revealing the privacy-sensitive part of the data.

- 1) This system will allow patients to upload their case history or reports on common servers from their homes and the doctors will review the reports and suggest treatment accordingly.
- 2) If the patients are satisfied they can continue with suggested treatment or submit the reports with added comments to other expertise.
- 3) Finally the suggested medicines will be directly delivered at postal address of the patient.

6 SYSTEM ARCHITECTURE

The fig.1 gives the system architecture of proposed system,

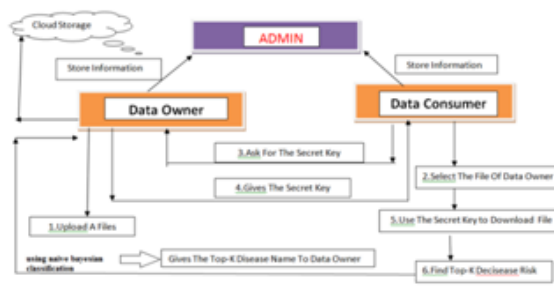


Fig. 1. System Architecture

The design and implementation of an architecture based on the combination of ontology, rules, web services, and the autonomic computing paradigm to manage data in home-based telemonitoring scenarios for the personal use.

- 1) Today every patient needs to be physically present in front of the doctor for OPD and considering the rate of growth of the population, it won't be possible for doctors to look after all patients.

- 2) Working people face issues of time for appointment mean their presence at the time when doctors present over there.
- 3) But we can provide solutions by giving working people facility of home surveillance.
- 4) Common patients cannot afford fees of highly qualified and renowned doctors. So suggestion will be given by the system to the patient.

7 MATHEMATICAL MODEL

Input :

- Data Owner(Z) = {D1,D2,D3,D4.....Dn}
- Data Consumer(Z) = {C1,C2,C3,C4....Cn}
- Admin = A
- F(Z) = {f1,f2,f3,f4.....fn }....file
- S(Z) = {s1,s2,s3,s4.....sn }....sec key
- P(Z) = {p1,p2,p3,p4.....pn }...public key
- MAC(Z) = {m1,m2,m3,m4.....mn }
- Download(Z) = { d1,d2,d3.....dn }
- Server(Z) = { S1,S2,S3...Sn }

Output:

Security of data with the help of Naive Bayesian Classification.

Mathematical Relation:

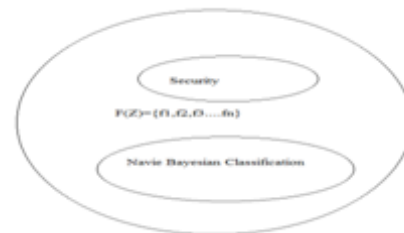


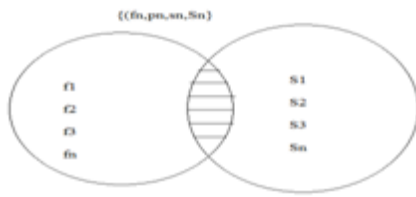
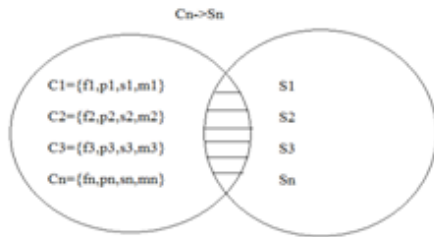
Fig. 2. F (Z) U Server (Z)

- Data Owner (Uploading):

- Data Consumer (Downloading):

Data owner is upload file respective consumer.

- Cn = {fn,pn,sn,Sn }....file

Fig. 3. $F(N) \cup s(N) \cup S(N)$ Fig. 4. $F(Z) \cup \text{Server}(Z)$

8 ALGORITHM

- 1) START.
- 2) Registration of Consumer.
- 3) Admin Accept Request and Gives the response.
- 4) Owner login.
- 5) Owner Allocate Space respective Consumer.
- 6) File Upload On Server for respective Consumer
- 7) To create Secrete key as well as master key.
- 8) Consumer login.
- 9) File Access or Disease Type search based on Naive Bayes Classifier.
- 10) download file=(secrete key, master key)
- 11) End

Naive Baye's Classification algorithm used for Top-K disease name retrieval:

- 1) First f classes select as follows,
- 2) $F(z) = \{c1, c2, c3, \dots, cn\}$
- 3) Take sample
- 4) $X(Z) = \{X1, X2, \dots, Xn\}, \dots, n$ -input sample
- 5) $A(Z) = \{A1, A2, A3, \dots, An\}, \dots, n$ value attribute
- 6) The classifier needs to predict X belongs to the class with the highest a posteriori probability, i.e., X is predicted to lie in

$$P(\vec{C_i} | X) > P(\vec{C_j} | X), \text{ for all } 1 \leq j \leq f, j \neq i.$$

$$P(\vec{C_i} | X) = \frac{P(X | C_i) P(C_i)}{P(X)}$$

$$P(X | C_i) \sim \prod_{k=1 \text{ to } n} P(X_k | C_i)$$

the class C_i if and only if there exists i , such that gives in Eq.3:

- 7) Apply Baye's theorem as in Eq. 4
- 8) To check neural category in the $P(X)$
- 9) sample value as $P(C1-X), P(C2-X), P(C3-X), \dots, P(Cn-X)$ calculated by Eq.5
- 10) Divide neural word from one site and non-neural word from one site.
- 11) Result

9 EXPECTED RESULT ANALYSIS

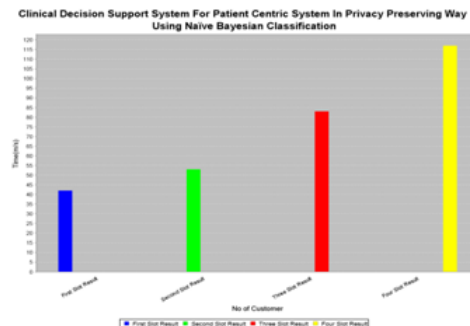


Fig. 5. Result Analysis

The Primitive of frequency mining is simple, but is very useful in data mining applications. Correspondingly, our privacy-preserving frequency mining solution is also quite simple, but is potentially useful whenever Server's Computation Time for a Single Frequency Calculation Privacy is a top concern. In this section, To demonstrate the power of our primitive by showing privacy preserving naive Bayes classifier computation in the fully distributed setting (which can be thought of

a horizontally partitioned database in which each record is held by a different party).

10 CONCLUSION AND FUTURE SCOPE

From the proposed system finally I conclude that the file can store or upload securely on server and access securely from the server. Also, classification of medical data through naive Bayesian classifier for efficient use of patient.

The Future Enhancement that a cryptographic approach that is efficient even in a many-customer setting, provides strong privacy for each customer, and does not lose any accuracy as the cost of privacy. Unlike general-purpose cryptographic protocols, this method requires no interaction between customers, and each customer only needs to send a single flow of communication to the data miner. However, we are still able to ensure that nothing about the sensitive data beyond the desired frequencies is revealed to the data miner.

- Alert System
- File Store on Server Replica Format.

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, IEEE Journal Of Biomedical And Health Informatics, Vol. Xx, No. Xx, December 2014"
- [2] "E. S. Berner, Clinical Decision Support Systems. Springer, 2007."
- [3] "M. A. Musen, B. Middleton, and R. A. Greenes, Clinical decision support systems, in biomedical informatics. Springer, 2014, pp. 643 674."
- [4] "H. Monkaresi, R. A. Calvo, and H. Yan, A machine learning approach to improve contactless heart rate monitoring using a webcam, IEEE J. Biomedical and Health Informatics, vol. 18, no. 4, pp. 11531160, 2014."
- [5] "C. Schurink, P. Lucas, I. Hoepelman, and M. Bonten, Computer assisted decision support for the diagnosis and treatment of infectious diseases in intensive care units, The Lancet infectious diseases, vol. 5, no. 5, pp. 305312, 2005."
- [6] "I. Kononenko, Machine learning for medical diagnosis: history, state of the art and perspective, Artificial Intelligence in medicine, vol. 23, no. 1, pp. 89109, 2001."
- [7] "N. Lavrac, I. Kononenko, E. Keravnou, M. Kukar, and B. Zupan, Intelligent data analysis for medical diagnosis: using machine learning and temporal abstraction, AI Communications, vol. 11, no. 3, pp. 191218, 1998."
- [8] "Y. Elmehdwi, B. K. Samanthula, and W. Jiang, Secure k-nearest neighbor query over encrypted data in outsourced environments, arXiv preprint arXiv: 1307.4824, 2013."
- [9] "K. M. Leung, Naive Bayesian classifier, Polytechnic University Department Of Computer Science/Finance and Risk Engineering, Copyright c, mleung@poly. edu, 2007."
- [10] "Transforming health care through big data strategies for leveraging big data in the health care industry, http://ihealthtran.com/wordpress/2013/03/iht_C2_B2-releases-bigdata-research-report-download-today/, 2013".
- [11] "Mr. P. A. Kharat, Dr. S. V. Dudul, IEEE 2011, Clinical Decision Support based on Jordan/Elman Network
- [12] "Mrudula Gudadhe, Kapil Wankhade and Snehlata Dongre, IEEE 2010, Decision Support System for heart disease using support vector machine and artificial neural network
- [13] "R. R. Janghel, Anupam Shukla and Ritu Tiwari, IEEE 2009, Clinical Decision Support System for Fetal Delivery using artificial neural network.
- [14] "AY Al-Hyari, A. M. Al-Tae and M. A. Al-Tae, IEEE 2013, Clinical Decision Support for diagnosis and management of Chronic Renal Failure."
- [15] "Abeer Y. Al-Hyari, Ahmad M. Al-Tae and Majid A. Al Tae, IEEE 2013, Clinical Decision Support System for Diagnosis and Management of Chronic Renal Failure



Ms. Manodnya A. Shitole Perusing masters in Computer Engineering at AVCOE, Sangamner. She has received her Bachelors in Computer Engineering from GESCOE'S, Nasik. She is a member of Association of Computer Machinery (ACM).



Prof. Manoj A. Wakchaure. Assistant Professor in department of Computer Engineering at AVCOE, Sangamner. He is a member of Association of Computer Machinery (ACM).