ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Implementing Clinical Decision Support System Using Support Vector Machine

Meenal V. Deshmukh, Prof. Pritish A. Tijare, Prof. Swapnil N. Sawalkar

M. E Student, Dept. of CSE, Sipna College of Engineering and Technology, Amravati, India

Associate Professor, Dept. of IT, Sipna College of Engineering and Technology, Amravati, India

Assistant Professor, Dept. of IT, Sipna College of Engineering and Technology, Amravati, India

ABSTRACT: To speed up the diagnosis time and improve the diagnosis accuracy in today's Healthcare system, it is important to provide a much cheaper and faster way for diagnosis. This system is called as 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 in recent years. The advantages of clinical decision support system include not only improving diagnosis accuracy but also reducing diagnosis time. In this paper, we have use the data mining technique name Support Vector Machine (SVM), which works as classifier that offered many advantages over the traditional methods of data mining and opens a new way for clinicians to predict patient's diseases. As the system is built on the cloud platform it is necessary to add some features that meets the security requirement. Specifically, with large amounts of data related to healthcare is generated every day, the classification can be utilized to excavate valuable information that improve clinical decision support system. The homomorphic encryption technique is useful for preserving the patient's privacy on the cloud. Here, by using the SVM classifier and the encryption technique and try to make the Clinical Decision Support System more helpful for providing information about some important deceases more accurately and efficiently.

KEYWORDS: Clinical Decision Support System, Privacy Preserving, Support Vector Machine, Homomorphic Encryption.

I. INTRODUCTION

Healthcare industry has the global scope to provide health services for patients, has never faced such a massive amounts of electronic data or experienced such a sharp growth rate of data today. However, if no appropriate technique is developed to find great potential economic values from large amount healthcare data, these data might not only become meaningless but also requires a large amount of space to store and manage. Over the past two decades, the miraculous evolution of data mining technique has imposed a major impact on the revolution of human's lifestyle by predicting behaviors and future trends on everything which can convert stored data into meaningful information. These techniques are well suitable for providing decision support in the healthcare system [1]. 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. 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.

Clinical Decision Support System has been defined as an "active knowledge systems", which use two or more items of patient's data to generate case specific advice [2]. This implies that a CDSS is simply a decision support system that is focused on using knowledge management in such a way to achieve clinical advice for patient care based on multiple items of patient's data. The main purpose of modern CDSS is to assist clinicians at the point of care. This means that clinicians interact with a CDSS to help for analyzing and reach a diagnosis based on patient data. Naive Bayesian classifier, one of the popular machine learning tools, has been widely used recently to predict various diseases in CDSS [1]. Despite its simplicity, it is more appropriate for medical diagnosis in healthcare than some sophisticated techniques. The CDSS with naive bayesian classifier has offered many advantages over the traditional healthcare systems and opens a new way for clinicians to predict patient's diseases. However, its flourish still hinges on understanding and managing the information security and privacy challenges, especially during the patient disease

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

decision phase [3]. One of the main challenges is how to keep patient's medical data away from unauthorized disclosure. The usage of medical data can be of interest for a large variety of healthcare stakeholders. For example, an online direct-to-consumer service provider offers individual risk prediction for patient's disease. Without good protection of patient's medical data, patient may feel afraid that his medical data will be leaked and abused, and refuse to provide his medical data to CDSS for diagnosis. Therefore, it is crucial to protect patient's medical data.

With increasing amounts of data being generated by businesses, healthcare industry and researchers there is a need for fast, accurate and robust algorithms for data analysis [5]. Improvements in databases technology, computing performance and artificial intelligence have contributed to the development of intelligent data analysis. The primary aim of data mining is to discover patterns in the data that lead to better understanding of the data generating process and to useful predictions. One recent technique that has been developed to address these issues is the support vector machine. The support vector machine has been developed as robust tool for classification and regression in noisy, complex domains. Support vector machine is a very strong and sophisticated machine learning algorithm especially when it comes to predictive analysis. In case of, Healthcare diagnosis system also it works best as like Naive Bayesian classifier and also support the mechanism for providing security to the sensitive medical data.

II. ARCHITECTURE OF PROPOSED SYSTEM

In this we are trying to improve the existing system using Clinical Decision Support System based on Support Vector Machine (SVM). We are using Support Vector Machine Data Mining classification technique for Clinical Decision Support System. The system will work faster and efficient using SVM [7]. It is widely used in real-life applications because of its simplicity and good performance both in theory and practice. However, in large-scale problems, where huge training data are available and must be used, such as road sign detection, the method's training and test phases might be prohibitively demanding in terms of computations. Thus, for large-scale problems the reduction of computational complexity is essential. We are using encryption techniques for preserving privacy of patient's data. And to preserve the privacy of tha data going through network we are using Homomorphic Encryption Technique to re-encrypt the data. All the processing will be done at server side and on the encrypted data.

We have defined the system model of CDSS in Fig.1, including Trusted Authority (TA), Cloud Platform (CP), Data Provider (DP), Processing Unit (PU), and Undiagnosed Patient (PA).

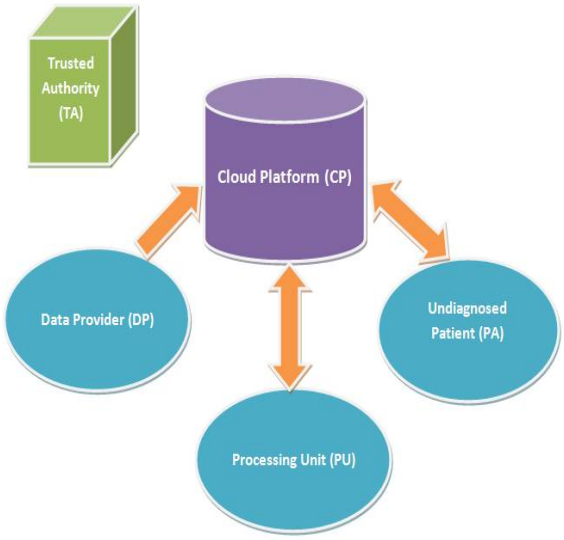


Fig.1: A System Model Under consideration

A. Trusted Authority (TA): TA is the indispensable entity which is trusted by all entities involved in the system, who is in charge of distributing and managing all private keys involved in the system.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

B. Cloud Platform (CP): CP contains unlimited storage space which can store and manage all the data in the system. Other parties who have limited storage space can outsourced their data to CP for storing.

C. Data Provider (DP): DP can provide historical medical data that contain patient's symptoms and confirmed diseases, which are used for training SVM classifier. All these data are outsourced to CP for storing.

D. Processing Unit (PU): PU can be a company or hospital which can provide online direct-to-customer service and offer individual risk prediction for various diseases based on client's symptoms. PU uses medical data to construct SVM classifier and then use the model to predict the disease risk of undiagnosed patients.

E. Undiagnosed Patient (PA): PA has some symptom information which is collected during doctor visits or directly provided by patient. (e.g. blood pressure, heart rate, weight, etc.). The symptoms can be sent to PU for disease diagnosis.

III. DESIGN GOALS

In order to achieve the secure medical decision for undiagnosed patient under the aforementioned model, our system design will fulfill privacy and performance guarantees as follows:

• **The proposed system should achieve privacy- preserving requirements**.
As stated above, if CDSS does not consider the privacy requirements, patient's highly sensitive information (symptom and disease information) will be disclosed to PU, CP, and unauthorized parties in the patient's medical decision. It will let patient unwillingly provide its own data to CDSS. In addition, PU is always a profit company which prevents his own data from leaking to other parties in the system. Therefore, the proposed system should achieve the privacy of PA and PU simultaneously.

• **The proposed system should achieve computation efficiency**.
The patient always has limited computational resources which cannot support overburden computation. To support patient-centric diagnosis results retrieval from CP in time, the proposed system should consider computation efficiency. Therefore, it is important to allow PA to retrieve diagnosis results in real time.

IV. IMPLEMENTATION STRATEGIES

A. Stepwise Work Flow of System:

Step 1: Undiagnosed patient will send his/her symptoms to the Cloud Platform (CP) in the encrypted format, using his/her public key.

Step 2: Data provider will provide the historical medical data to the CP in encrypted format using homomorphic encryption technique.

Step 3: The CP will decrypt this data and sends to SVM classifier for training. Once the training will be done the disease risk will be calculated based on the symptoms provided by the undiagnosed patient and the training result. All the processing is done on encrypted data, which preserves the privacy of patient's data.

Step 4: Once the disease risk will be calculated, the predicted result will be send to the next level. At this level the probability of predicted disease risk will be calculated and according to patients preferences the results will send to the patient in encrypted form.

Step 5: If the patient wants top-k predicted disease names then they can give their own preferences accordingly. For this, at the server side we will be using top-k algorithm. In this algorithm the maximum probability disease risk will be calculated. And the top-k results will be sent to the patients as per his/her preferences. Once the encrypted diagnosis result will get reached at the client side, the undiagnosed patient will decrypt these results by using his/her private key.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

B. Basic flow of proposed system

Here, we have use Homomorphic encryption technique to provide security on cloud. Homomorphic encryption is a new concept of security which enables providing results of calculations on encrypted data without knowing the raw data on which the calculation was carried out, with respect of the data confidentiality. We have also used the RSA and Paillier algorithm for homomorphic encryption using proxy Re-encryption algorithm that prevents cipher data from Chosen Ciphertext Attack (CCA). So this system is more secure than existing system. The flowchart containing complete flow of our proposed system is shown below.

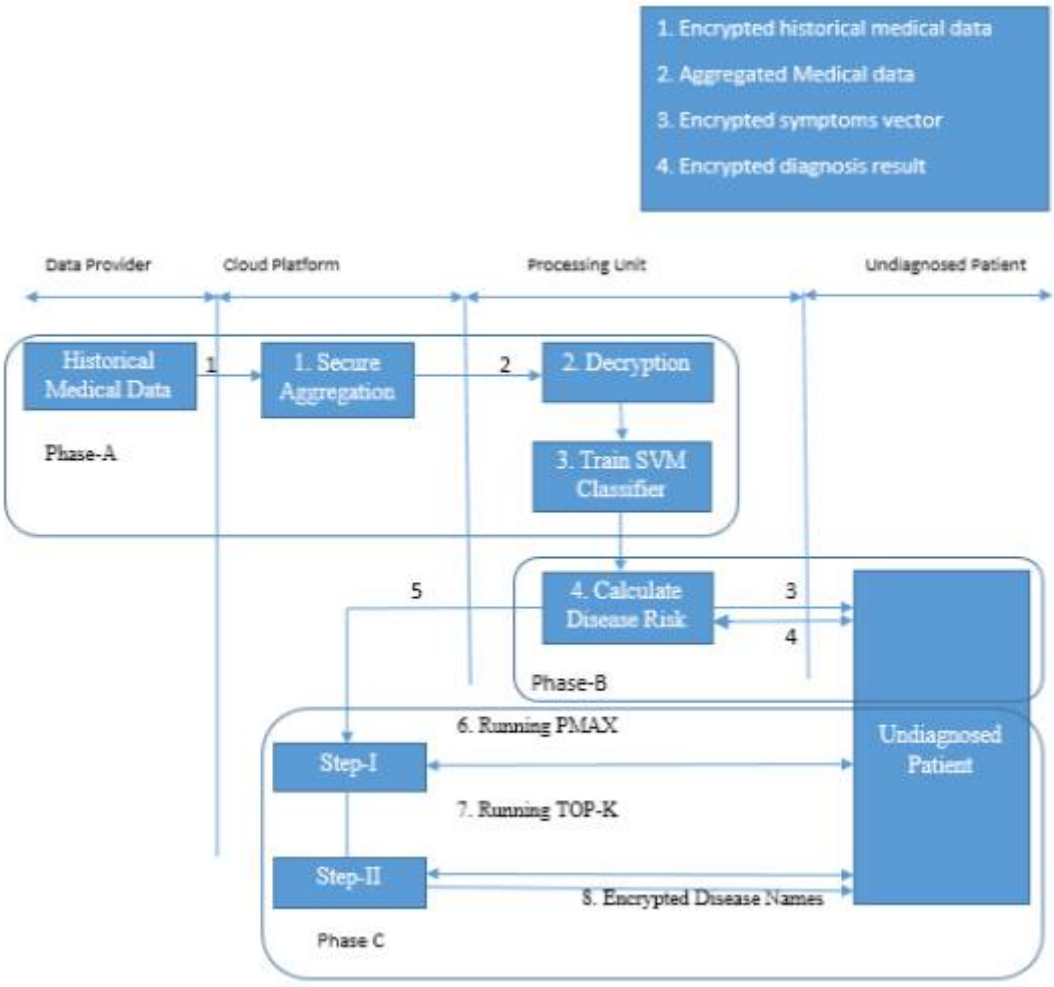


Fig. 2: Basic flow diagram for execution of proposed system.

The execution of proposed system is given by following steps:

- Client will request for Public key to the key store.
- Key store will provide public key to the client as per his / her request.
- Client will provide their symptoms for proper diagnosis.
- Provided symptoms will get encrypt using RSA algorithm.
- Then symptom .arff file i.e. (attribute relation file format) will get created.
- Historical Medical Data (dataset file) will get encrypted using symmetric key (DES algorithm).

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

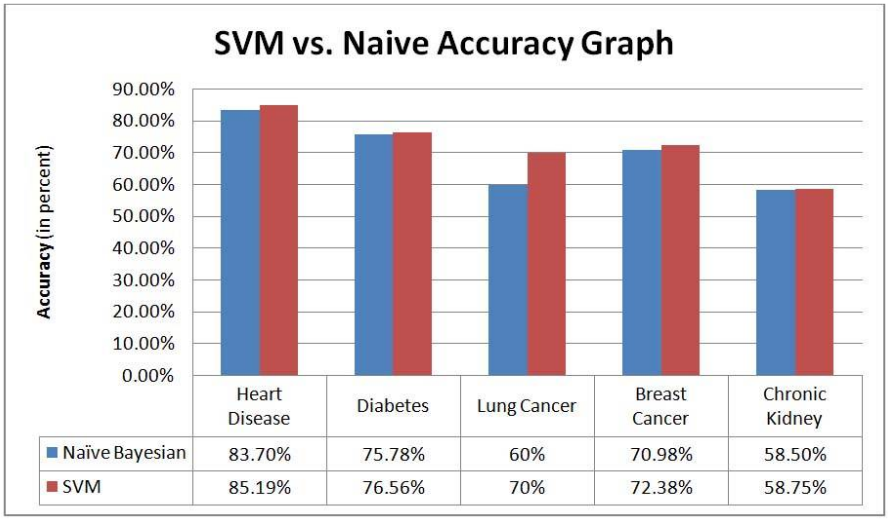
(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

- The server will read both encrypted input symptoms file and dataset file.
- Paillier encryption will performed on both files.
- Then both encrypted files will send to SVM classifier for training and testing purpose.
- After training and testing process, SVM will provide prediction calculation and calculates summary.
- For Sending result to Client, SVM performs decryption on prediction data using Pailliers key.
- Client will decrypt the result obtained using his/her private key.
- Client will observe the predictions.

V. EXPERIMENTAL RESULT AND DISCUSSION

The basic flow of execution of proposed methodology is shown in figure 2 above. Step-by-step workflow of our system is also explain above. Experimental result shows that Clinical Decision Support System Based on Support Vector Machine gives better prediction result than naïve bayes classifier. We have taken the example of for four most important disease i.e. cancer disease (Lung cancer, Breast cancer), heart disease, diabetes, and chronic kidney. In terms of accuracy, when we compare the accuracy of result generated using SVM classifier, it is find out that predictions generated by SVM Classifier is 5% to 10% more accurate than Naive bias algorithm. The comparison graph of the accuracy for finding out prediction between naïve Bayes and our SVM classifier is shown in the Graph 1. It is found out that SVM classifier gives prediction for all types of diseases are is 5 to 10 present more accurate than naïve Bayes algorithm.



Graph 1: Comparing accuracy of prediction between SVM and Naïve Bayes

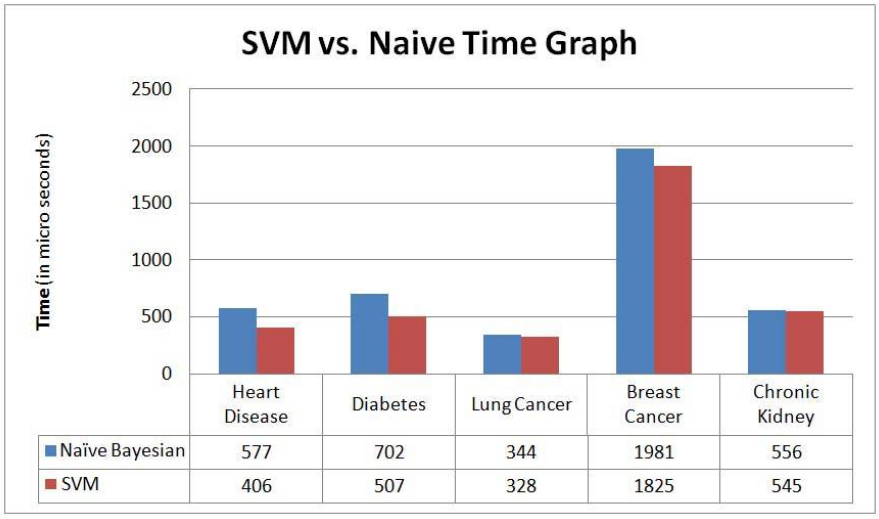
In terms of Time efficiency for finding out prediction of mostly chronic diseases traditional naïve Bayes algorithm takes more time as compared to our proposed algorithm of support vector machine (SVM). The graph statics of comparison for time efficiency between SVM and Naïve Bayes is shown below (Graph 2). We will find out that SVM is more time efficient and generated more accurate results than traditional naïve Bayes algorithm.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016



Graph 2: Comparing Time Efficiency of prediction between SVM and Naïve Bayes

Advantages:

By designing the system like this we are able to

- Improving diagnosis accuracy for any critical diseases
- Reducing diagnosis time gives proper prescription in much less time
- High disease prediction success rate without any kind of burden
- Reducing communication overhead
- Preserving privacy of patient's data

VI. CONCLUSION

In this paper, we have proposed Clinical decision support system using the classification technique o data mining called Support Vector Machine. Using SVM, the computational time and diagnosis rate in our system gets improved. SVM has excellent performance in generalization so it can produce high accuracy in classification for diagnosis. The patient can securely retrieve top-k diagnosis result according to their own preferences. With the advantage of Homomorphic encryption technique, the patient's privacy over the cloud will achieved. The processing is done on the encrypted data, so that there is no loss in the privacy of patient's data while training the SVM classifier. These results evidentially proved the proposed method and showing the nice performance of classification accuracy based on dataset of Diabetes. From the implementation system we can conclude that the results of proposed work are better than result of previous work as the classification accuracy of SVM classifier is 76.56% which is more than the accuracy of Naïve bayes that is 75.78%.

REFERENCES

[1] Ximeng Liu, Rongxing Lu, Jianfeng Ma, Le Chen, and Baodong Qin, "Privacy- Preserving Patient-Centric Clinical Decision Support System on Naïve Bayesian Classification", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. XX, NO. XX, DECEMBER 2014.
[2] R. S. Ledley and L. B. Lusted, "Reasoning foundations of medical diagnosis," Science, vol. 130, no. 3366, pp. 9–21, 1959.
[3] H. R. Warner, A. F. Toronto, L. G. Veasey, and R. Stephenson, "A mathematical approach to medical diagnosis: application to congenital heart disease," Jama, vol. 177, no. 3, pp. 177–183, 1961.
[4] 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. 305–312, 2005.
[5] M. Kantarcıoglu, J. Vaidya, and C. Clifton, "Privacy preserving naive bayes classifier for horizontally partitioned data," in IEEE ICDM work-shop on privacy preserving data mining, 2003, pp. 3–9.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

[6] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu, "Tools for privacy preserving distributed data mining," ACM SIGKDD Explorations Newsletter, vol. 4, no. 2, pp. 28–34, 2002.
[7] X. Yi and Y. Zhang, "Privacy-preserving naive bayes classification on distributed data via semi-trusted mixers," Information Systems, vol. 34, no. 3, pp. 371–380, 2009.
[8] A. Amirbekyan and V. Estivill-Castro, "A new efficient privacy- preserving scalar product protocol," in Proceedings of the sixth Aus- tralasian conference on Data mining and analytics-Volume 70. Aus- tralian Computer Society, Inc., 2007, pp. 209–214.
[9] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," IEEE Network, vol. 28, no. 4, pp. 46–50, 2014.