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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

A Survey towards Patient Centric Clinical Decision Support System Using Navie Bayes Classification System

Trupti . S.Mokati, Prof. Vijay B. Gadicha

M.E Student, Dept. of CSE, P.R.Pote (Patil) College of Engineering and Management, Amravati, India

HOD, Dept. of CSE, P.R.Pote College, Amravati, India

ABSTRACT: Clinical Decision Support System (CDSS), assist the physicians in diagnosing patient disease with similar symptoms, and suggest proper treatment for them in much more efficient manner. The advantage of clinical diagnosis support system is that it reduce the diagnosis time and give accurate prescription to all the diseases available in our database. This paper performs the study on different data mining techniques like Naive Bayesian classifier, SVM and many more that are available for the CDSS that helps to offered many advantages over the traditional healthcare systems and opens a new way for clinicians to predict patient's diseases. As healthcare is the field in which Security of patient's data is paramount important, this paper also provide the study of different security mechanisms that are suitable for the CDSS along with data mining techniques. The security mechanisms that is RSA and Holomorphic encryption technique seems more proper that meets the security Goals for CDSS. As the large amounts of clinical data generated every day, our data mining technique that mainly require for classification should be compatible enough that can be utilized to execute valuable information to improve clinical decision support system. The paper shows that as the clinical data mainly related to the patients are growing on increasing and reached towards storage on cloud. So now to extract important information from this huge amount of data securely, requires Homomorphic encryption techniques and SVM classification techniques that results in generation of proper Clinical Decision Support System.

KEYWORDS: Clinical Decision Support System (CDSS), Data Mining, Privacy Preserving, Naive Bayesian classifier, Support Vector Machine, Homomorphic Encryption.

I. INTRODUCTION

Over the past two decades, Health care industry abundantly broadcast in the global scope to provide health care and health care services to patients', has never faced such a such a growth in technological side. However, if 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. Different Data mining that are available has more potential for the healthcare industry to enable health systems to automatically analyze and provide security to the historic data which is stored in cloud. Over the past few years, there massive improvement in data mining technique has a major impact on human's lifestyle by predicting behaviors and future trends. To reduce the diagnosis time and to improve the accuracy, a new diagnosis system should be developed to provide diagnosis in a faster way.

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. 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 to analyses, and reach a diagnosis based on patient data. Naive Bayesian classifier, one of the popular machine learning tool of data mining, 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

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

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.

The patient medical data is of large amount along with this one of the main challenges is how to keep patient's medical data away from unauthorized disclosure. Various approaches, such as cryptographic and non-cryptographic are used to preserve the privacy of health-data in the cloud. The majority of the solutions use certain cryptographic techniques to conceal the contents of health records, while quite a few solutions are based on non-cryptographic approaches using policy based authorizations. The benefit of cryptographic techniques is that they not only are capable of encrypting the data in storage and over the network [2], but also employ authentication mechanisms requiring decryption keys and verification through digital signatures. Moreover, fine-grained and patient-centric access control mechanisms have also been deployed that enable patients to specify the individuals who could have access to health-data [3]. Furthermore, quite a few privacy preserving solutions allow the patients themselves to encrypt the health data and provide the decryption keys to the individuals with right-to-know privilege. This chapter encompasses the recent efforts that have been made to preserve the privacy of the health-data in the cloud environment.

In this paper, we have studied different mining techniques used by different researchers for securely mining and storing large amount of data from clinical decision support system (CDSS). The different data mining algorithms like Naïve bayse, Neural Networks, Support Vector Machine (SVM), etc are also studied to findout one of the most powerful classification techniques that was successfully applied to real world problems of CDSS. This has greater advantage in case of improving diagnosis accuracy in clinical decision support system. Along with this, we have studied the RSA and homomorphic Encryption technique for providing security to the sensitive data related to patient health information. The remaining paper is organized as, Section II gives some Literature Survey which gives brief information of the study done by different researchers in the field of Clinical Decision Support System CDSS. Section III discuss in brief about some the data mining and Privacy preserving technique used for providing CDSS. Finally, Section IV concludes the paper.

II. LITERATURE SURVEY

The authors, Ximeng Liu, Rongxing Lu, Jianfeng Ma in [1] proposed a privacy-preserving patient-centric clinical decision support system using naïve Bayesian classifier. By taking the advantage of emerging cloud computing technique, processing unit can use big medical dataset stored in cloud platform to train naïve Bayesian classifier. And then apply the classifier for disease diagnosis without compromising the privacy of data provider. But in this system, the patient can securely retrieve the diagnosis results according to their own preference entered in the system. For the security mechanism authors provide all the data are processed in the encrypted form, that helps to achieve patient-centric diagnose result retrieval in privacy preserving way. But as the data is growing on increasing in much more faster way that reaches to cloud, one has to use more efficient data mining technique that helps in privacy preserving patient-centric clinical decision support systems.

The authors V. Krishnaiah, G. Narsimha, N. Subhash Chandra [4] in this paper, gives study of different data mining techniques that can be employed in robotic heart disease prediction systems. The analysis shows that different technologies are used in all the papers with taking different number of attributes reached their results different accuracy depends on tools used for implementation. Even though applying data mining techniques to assist health care professionals in the diagnosis of heart disease is having various successes. The symbolic Fuzzy K-NN classifier can be tested with the unstructured data available in health care industry data base by modifying into fuzzified structured data with increased attributes and with a collection of more number of records to provide better accuracy to the system in predicting and diagnosing the patients of heart disease. Authors here provides a fast and simple thoughtful of different prediction models in data mining and helps to find greatest model for further work. But at the same time, this work can be enhanced by increasing the number of attributes for the existing system of our previous work.

The authors Luis Tabares, Jhonatan Hernandez, Ivan Cabezas [5] in this paper, said that Cloud computing has proved to be a feasible solution to currently growing healthcare sector. As all the needy once required healthcare services it should be cost-effective, ubiquitous and elastic model, enabling shared computing resources between healthcare providers and patients and that is possible by cloud computing platform. They suggested that three main components

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

such as a knowledge base, an inference engine or an artificial intelligence component, are required to expose the knowledge by means of a service would be a convenient approach to implement alerts and reminders, knowledge service and diagnostic/treatment CDSS, and apply them properly. Findings in this paper said that several authors may be not properly using the terms cloud-based and cloud computing, since they are focusing on service-based or web-based architectures and also they are not detailing the conducted architectural design process. This will raises concerns and worries, since such lack of rigor on the software engineering process does not allow identifying considered sources and used methods for gathering quality scenarios. As there was not any application of architectural evaluation methods based on scenarios. Consequently, well known concerns such as security and privacy may not be being well validated in practice.

The authors Shreya Anand, Ravindra B Patil, Krishnamoorthy P [6] in this paper [], provides some insights on the different risk models available for assessment of Cardiovascular disease (CVD) risk. Here, the strength and limitations of each of these models are found that there is no India specific CVD risk. To assist the primary care physicians for early diagnosis and management of chronic diseases such as CVDs, author develop WHO/ISH based risk stratification model based android based application. The application develop can be installed in the hospital setup or on top of EMR solution to provide risk stratification as well as life style and medication recommendations to the subjects. This solution helps less well trained primary care physicians in the diagnosis and management of chronic diseases such as CVDs, assisted by clinical decision support systems (CDSSs). But as this solution is derived mostly for chronic diseases so it is necessary to develop the similar approach for other emerging countries for better delivery of healthcare.

The authors Kulwinder Singh Mann, Navjot Kaur [7] in this paper, designed and develop the Framework for mining big data using Mapreduce and Hadoop mainly for predicting Heart diseases accurately with reduced attributes. Inference rules designed here for building Knowledge based of CDSS are generated by traversing nodes of accurate Mapreduce are based decision trees. These trees are generated by classification via clustering approach and are more accurate as compared to WEKA's decision trees. They added a Combiner between Mapper and Reducer to improve the performance of C4.5 decision tree algorithm. Using reduced attributes instances are more accurately classified with classification via clustering approach. They have usedthe Hadoop clusters for deploying on GCS using GCS connector. The Mapreduce programming is simplified as it eliminate the need of accessing HDFS for data files. Due to Global replication reliability of system is increased.

The authors Jussi Mattila, Juha Koikkalainen, Arho Virkki, Mark van Gils, *Member IEEE*, and Jyrki L¨otj¨onen [8] in this paper, gives the design and implementation of a generic decision support system was presented. Here a reusable software library employing a statistical disease state modeling method is used, which is able to robustly analyze heterogeneous multiscale patient data with minimal preprocessing. The method context-agnostic data access, analysis, and visualization used here by authors allow the library to be rapidly applied in several contexts. As long as definitions of the data and the data itself are provided to the library, it can organize available values and construct interactive views that provide analyses of the recently defined information to clinical decision makers. Here the goal of the author is to provide evidence-based decision support for clinicians during diagnostic work. Application of the decision support library was demonstrated by developing a prototype CDSS tool for early prediction of AD. Author apply DSI method and the decision support library to several other datasets to assess their robustness more comprehensively.

The author Schurink et al. [9] discuss the computer-based decision-support systems to assist Intensive Care Unit (ICU) physicians in the management of infectious diseases. In this paper, author described several computer models such as bayesian networks that may be used in clinical practice in the near future. But at the same time, as the privacy of the patient's information becomes more and more important, naive Bayesian classification were considered as a challenge to privacy-preservation due to their natural tendency to use sensitive information about individuals.

The author Chuang et. al [10] revealed the means of effectively using a number of validation sets obtained from the original training data to improve the performance of a classifier. The proposed validation boosting algorithm by author was illustrated with a support vector machine (SVM) in Lymphography classification. A number of runs with the algorithm was generated to show its robustness as well as to generate consensus results. At each run, a number of

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

validation datasets were generated by randomly picking a portion of the original training dataset. The results showed that the proposed method with validation boosting could achieve much better generalization performance on repeated iterations with a testing set than the case without validation boosting.

The authors Jos´e A. Reyes-Ortiz, Beatriz A. Gonz´alez-Beltr´an, Lizbeth Gallardo-L´opez [11] in his work, introduced a detailed analysis of existing approaches to clinical decision support using NLP that is Natural Language Processing. They help CDSS by providing a solution based on Natural Language Processing from free texts. Authors aims to perform an information extraction or chunk text classification for this they have performed an analysis based on several features such as: language, free-text, NLP approach, patients, and Clinical decision support task and health outcomes. They become the present work as a valuable analysis. The approaches perform an information extraction from free text. From this some significant issues and open problems are generated as, the complexity of natural language differs between languages like English, Spanish, German semantic and also there is a difference between semantic and morphological components of different languages. Another challenges to NLP is to involve clinical decision support for involving semantic and linguistic methods with good accuracy.

III. TECHNOLOGIES USED FOR DESIGNING

3.1 Data mining techniques

Data mining techniques have been widely used in clinical decision support systems (CDSS) that performs prediction and diagnosis of various diseases with better accuracy. The techniques have been very effective and helps in developing clinical support systems because they are able to detect hidden patterns and relationships in medical data. There are large no. of classification techniques which can be used for clinical decision support system. The aim of classification is to predict the target class for each case in the data accurately. Classification is important when a repository of data contains samples that can be used as the basis for future decision making. Some of the data mining that are mainly used for classification in CDSS are given below:

A. Naive Bayes Classifier

Naive Bayes have been widely used in data mining for data classification [12]. They have a high generalization ability which provides high reliability in real-world applications like image processing, computer vision, text mining, natural language processing, biomedical engineering, and many more [12]. The Naive Bayes is to separate classes by a classification function, which is obtained by training with the data samples. Depending on the precise nature of the extent to which something is likely to happen using the naive bayes classifiers can be trained very efficiently in a supervised learning approach. In most appliances naive Bayes models uses the method of maximum likelihood; simply work with the naive Bayes model without believing in Bayesian probability or using any Bayesian methods.

B. Neural Network

The original development of the neural network programs was inspired by the way the brain recognizes patterns [13]. Neural Networks is allows the systems to learn from existing knowledge and experiences. The three main layers of Neural Networks are Input, Output and Hidden layer. Neural Network is made of nodes that is called neurons. And there is weighted connection between nodes of different layers, which is used to transfer signals between the nodes. In supervised neural networks, examples in the form of the training data are provided to the network one at a time. Neural Network can continue with incomplete data that may be come in clinical data support to give educated guesses about missing data and get improved with every use due to its adaptive system learning.

C. Decision Tree

The use of decision trees is perhaps the easiest to understand and the most widely used method that falls into the category of supervised learning. A typical decision tree system adopts a top-down strategy in searching for a solution. It consists of nodes where predictor attributes are tested. At each node, the algorithm examines all attributes and all values of each attribute with respect to determining the attribute and a value of the attribute that will “best” separate the data into more homogeneous subgroups with respect to the target variable. In other words, each node is a classification

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

question and the branches of the tree are partitions of the data set into different classes. This process repeats itself in a recursive, iterative manner until no further separation of the data is feasible or a single classification can be applied to each member of the derived subgroups. AY AI-Hyari et al 2013 developed a CDSS for diagnosing patients with Chronic Renal Failure using different classification methods like neural network, naïve bays and decision tree. They proved that there is (92.2%) accuracy of using Decision tree algorithm as compared to all other algorithms [14]. They applied supervised decision tree classifier C4.5 to classify image samples with sensitivity of 98.1% and specificity of 99.6%.

D. Support Vector Machine

Support vector machine (SVM) has become more and more popular tool in task of machine learning involving classification, regression etc. Support Vector Machine (SVM) is primarily a classier method that performs classification tasks by constructing hyper planes in a multidimensional space that separates cases of different class labels [15]. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. In machine learning, support vector machines analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. Support Vector Machine is a state of the art classification. It performs well with real world application such as classifying text, classifying images etc. SVM are the standard tools for machine learning and data mining. And with this large amount of application and advantage, we will also use SVM for our proposed classification technique in CDSS.

3.2 Privacy Features

Privacy is crucial for the success of patient's diseases diagnosis. An algorithm which utilizes the healthcare knowledge available in the remote location via the Internet while preserving privacy needs to be develop. The clinician encrypts each element of the patient data using the public key and sends the encrypted data and the corresponding public key to the server. Due to the encryption is performed with the clinician's public key, no one including the server could decrypt this to obtain the values of the elements. Thus, the patient data are protected against being revealed even to the server taking part in this process. Since the server only has the encrypted patient data, it has to compute in the encrypted domain using homomorphic and two-party secure computation properties

A. RSA Encryption

RSA is one of the first practical public-key cryptosystems and is widely used for secure data transmission. In such a cryptosystem, the encryption key is public and differs from the decryption key which is kept secret [16]. In RSA, this asymmetry is based on the practical difficulty of factoring the product of two large prime numbers, the factoring problem. With the property of encryption and decryption provided by RSA, one can use it for perform encryption of Patients sensitive information for storage and some where large database mainly cloud. With the facility of decryption with private key on can download particular person's data securely by giving his private keys. RSA passes encrypted shared keys for symmetric key cryptography which in turn can perform bulk encryption-decryption operations at much higher speed. This features is most useful as the data available for this in clinical decision support system CDSS is of large amount mostly stored on cloud.

B. Homomophic Encryption

Homomorphic encryption is a form of encryption that allows computations to be carried out on ciphertext, thus generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext [17]. This is sometimes a desirable feature in modern communication system architectures. Homomorphic encryption would allow the chaining together of different services without exposing the data to each of those services. For example, a chain of different services from different companies could calculate 1) the tax 2) the currency exchange rate 3) shipping, on a transaction without exposing the unencrypted data to each of those services. Homomorphic encryption schemes are malleable by design. This enables their use in Clinical Decision Support System for ensuring the confidentiality of patient's data. In addition the homomorphic property of the cryptosystems can be used to create many other secure systems that needs for data related to Healthcare 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)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

IV. CONCLUSION

For speeding 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. This paper provides the study of different types of work done previously by different authors. There are several techniques studied for data mining such as Naive Bayesian classifier, SVM, etc. Out of which by selecting one of the popular machine learning tools, that has been widely used to predict various diseases in CDSS. Despite its simplicity, it is more appropriate for medical diagnosis in healthcare than some sophisticated techniques. The advantages of clinical decision support system include not only improving diagnosis accuracy but also reducing diagnosis time. For the purpose of providing security to the sensitive information of patient data some security techniques are studied. With the use of RSA encryption and Homomorphic encryption technique, the patient's privacy over the large amount of data will achieved. The data that will be transfer is present in encrypted data, so that there will be no loss in the privacy of patients data while training the classifier and providing data on the network.

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] *Assad Abbas, Samee U. Khan,* "e-Health Cloud: Privacy Concerns and Mitigation Strategies", *North Dakota State University*, USA.
[3] Jim Basilakis, Bahman Javadi, Anthony Maeder, "The Potential for Machine Learning Analysis over Encrypted Data in Cloud-based Clinical Decision Support – Background and Review", *Proceedings of the 8th Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2015)*, Sydney, Australia, 27 - 30 January 2015.
[4] V. Krishnaiah, G. Narsimha, N. Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review", *International Journal of Computer Applications (0975 – 8887)*, Volume 136 – No.2, February 2016.
[5] Luis Tabares, Jhonatan Hernandez, Ivan Cabezas member IEEE, "Architectural approaches for implementing Clinical Decision Support Systems in Cloud: A Systematic Review", *First Conference on Connected Health: Applications, Systems and Engineering Technologies*, 978-1-5090-0943-5/16 $25.00 © 2016 IEEE.
[6] Shreya Anand, Ravindra B Patil, Krishnamoorthy P, "An Analytics based Clinical Decision Support System for CVD Risk Assessment and Management", 978-1-4577-0220-4/16/$31.00 ©2016 IEEE.
[7] Kulwinder Singh Mann, Navjot Kaur, "Cloud-deployable health data mining using secured framework for Clinical decision support system", 978-1-4799-6908-1/15/$31.00 ©2015 IEEE.
[8] Jussi Mattila, Juha Koikkalainen, Arho Virkki, Mark van Gils, *Member, IEEE*, and Jyrki L¨otj¨onen, "Design and Application of a Generic Clinical Decision Support System for Multiscale Data", *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, VOL. 59, NO. 1, JANUARY 2012.
[9] C. Schurink, P. Lucas, I. Hoepelman, and M. Bonten, "Computer- assisted decision support for the diagnosis and treatment of infectious disease s in intensive care units," The Lancet infectious diseases, vol. 5, no. 5, pp. 305–312, 2005.
[10] Tzu-cheng Chuang, Okan K. Ersoy, Saul B. Gelfand, Boosting Classification Accuracy With Samples Chosen From A Validation Set, ANNIE (2007), Intelligent ` Engineering systems through artificial neural networks, St. Louis, MO, pp. 455-461.
[11] Jos´e A. Reyes-Ortiz, Beatriz A. Gonz´alez-Beltr´an, Lizbeth Gallardo-L´opez, "Clinical Decision Support Systems: A survey of NLP-based Approaches from Unstructured Data", 26th International Workshop on Database and Expert Systems Applications, 1529-4188/15 $31.00 © 2015 IEEE.
[12] Rahulamathavan, Y. Veluru, S. Phan, R.C.-W. Chambers, J.A. Rajarajan, M. ―Privacy-Preserving Clinical Decision Support System Using Gaussian Kernel-Based Classification‖, *IEEE Journal of BioMedical And Health Informatics*, Volume 18,No. 1,pp:56-66,2014.
[13] J. Michael Hardin and David C. Chhieng, "Data Mining and Clinical Decision Support Systems" [Online]available: http://cs.engr.uky.edu/~goldsmit/cdm/DMCDSS.pdf
[14]AY AI-Hyari, A. M. Al-Taee and M. A. Al-Taee, IEEE 2013, "Clinical Decision Support for diagnosis and management of Chronic Renal Failure.
[15] Support Vector Machine: Introductory Overview [Online] available: http://www.statsoft.com/Textbook/Support-Vector-Machines
[16] Smart, Nigel "Dr Clifford Cocks CB". *Bristol University*. Retrieved August 14, 2011.
[17] *Cortes, C.*; Vapnik, V., "Support-vector networks". *Machine Learning*. 20 (3): 273–297. doi:10.1007/BF00994018.