IJTIMES
International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017),e-ISSN:2455-2585
International Conference on
Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume-5, Special Issue-March, 2019.

Privacy-Preserving Patient-Centric Clinical Decision Support System on SVM

S Ragavi Priya[1],B.A.Lathika [2] M.Tech

[1] PG Scholar, Department of Computer Science & Engineering
[2] Assistant professor, Department of Information Technology
[1&2] KGiSL Institute of Technology, Coimbatore

Abstract—A safe decision support estimation in health care system preserves the privacy of the patient data, the decision estimation and the server side clinical support system. Clinical decision support system, which uses data mining technique to help clinician make proper decisions, has received considerable attention nowadays. The advantages of clinical decision support system include not only improving diagnosis accuracy it also reducing diagnosis time. With rapid growth of clinical data generated day by day, the classification techniques can be utilized to hollow out valuable information to improve a clinical decision support system. In this study, proposed a new privacy-preserving patient-centric clinical decision support system, which helps clinician complementary to diagnose the risk of patients' disease in a confidentiality way. The past patients' treatment history of the diseases are kept in the cloud environment which can be used to prepare the SVM (Support Vector Machine) classifier without disclose any human being health data, and then the trained classifier can be applied to compute the disease risk for new coming patients and also allow these patients to retrieve the top-k disease names according to their own preferences. In addition to that, to protect the privacy of past patients' treatment data, a classical cryptographic methodology called TDES (Triple Data Encryption Standard) scheme is proposed. The accuracy of the proposed system is evaluated via widespread simulations and also it demonstrates efficiency of patient's disease risk classification.

Keywords—Clinical Decision Support System, Support Vector Machine, Triple Data Encryption Standard

I. INTRODUCTION

A. Data Privacy

Data mining is the process of discovering new patterns from large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems. The goal of data mining is to extract knowledge from a data set in a human-understandable structure and involves database and data management, data pre-processing, model and inference considerations. Privacy has become an increasingly plays an vital role in data mining. Privacy concerns strictly restrict the free flow of information. Privacy is one of the most important properties of information. The protection of sensible information has a major role in Organization for a commercial do not want to reveal their private database and information. Even the individual does not want to reveal their personal medical data to other than those they give permission to. This implies that revealing of an instance of a particular class to be classified may be equivalent to revealing secret and private information.

The protection of sensible information has a relevant role. Privacy preserving data mining algorithms have been recently introduced with the aim of preventing the discovery of sensible information. The data id is generated in the healthcare system time by time so every patient is provided their personal information to the doctor for making the decision but the privacy is the major issue for healthcare system of patient. The main requirement is to provide the security to the patient data from unauthorized access, the privacy preserving clinical decision support system is given in the system.

B. Clinical Decision Support

Healthcare business, massively distributed within the international scope to produce health services for sick people, has never faced such a colossal amounts of electronic knowledge or old such a pointy rate of information these days. As declared by the Institute for Health Technology Transformation, health care knowledge alone reached one hundred fifty exabytes and would before long reach zettabyte scale and even yottabytes within the future. However, if no applicable technique is developed to search out nice potential economic values from huge attention knowledge, these knowledge won't solely become vacuous however additionally need an out sized quantity of area to store and manage.

Over the past two decades, the miraculous evolution of information mining technique has obligatory a significant impact on the revolution of human's style by predicting behaviors and future trends on everything, which might convert keep knowledge into meaningful information. These techniques are well suitable for providing decision support within the healthcare industry. To reduce the diagnosis time and improve the diagnosis accuracy, a replacement system in health care business ought to be practicable to produce means a far cheaper and quicker way for diagnosis. Clinical decision support system (CDSS), with data mining techniques being applied to help doctors/clinicians in identification of patient diseases with similar symptoms.

II. OVERVIEW OF THE PROJECT

Clinical decision system is the system used to diagnosis the patient disease in different way. It mainly checks the patient symptoms, test result and historical data. To predict the disease or treatment there are many more algorithms used based on medical phenomenon. To speed up the diagnosis time and improve the diagnosis accuracy, a new system in healthcare industry should be workable to provide a much cheapest and fastest way for diagnosis. Classical support vector machine classifier, one of the popular machine learning tools, has been widely used recently to predict various diseases in Clinical Decision Support System (CDSS). It performs well in multi class prediction.

The past patient's historical data are stored in cloud and can be used to train the support vector machine classifier without leaking any patient's historical /medical data and then the trained classifier can be applied to calculate the disease risk for new coming patients and also, allow these patients to retrieve the top-k disease names according to their own preferences.
This proposed system is to improve the performance of the clinical decision support system and to improve the diagnosis time and accuracy. To protect the privacy of past patients' treatment data, a classical cryptographic methodology called TDES (Triple Data Encryption Standard) scheme is proposed.

III. SYSTEM STUDY

A. EXISTING SYSTEM

In classification the naïve bayes classifier minimizes the probability of misclassification. Naïve Bayes have been widely used in machine learning for data classification. They have a high generalization ability which provides high reliability in real-world applications such as image processing, computer vision, text mining, natural language processing, biomedical engineering, and many more.The goal of an naive bayes is to separate classes by a classification function, which is obtained by training with the data samples.
In the existing system, Naive Bayes uses the kernel estimator for numeric attributes rather than a normal distribution and utilized supervised discretization while converting numeric attributes to normal. Bayesian classifier could represent the probabilistic relationships between diseases and symptoms. Naive Bayesian classifier is a classifier which has been proved to be effective in many practical applications, including text classification, medical diagnosis, and systems performance management.

Drawbacks

- If categorical variable has a category (in test dataset), which was not observed in training data set, then model will assign a 0 (zero) probability and will be unable to make a prediction on missing data.
- The limitation of Naïve Bayes is the assumption of independent predictors. In real life, it is impossible to get a set of predictors which are completely independent.
- The Naive Bayes is also known as a bad estimator, so the probability outputs from prediction are not to be taken too seriously.

B. PROPOSED SYSTEM

In this study proposed 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 earlier period data are stored in cloud and it can be used for SVM (Support Vector Machine) 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. Detailed privacy analysis ensures that patients' information is private and will not be leaked out during the disease diagnosis phase.
Support vector machine (SVM) has become more familiar tool for machine learning tasks involving classification, regression etc. SVM separate the data into two of performing classification and constructing an N-dimensional hyper plane. SVM is supervised learning model that is applied for classification.

SVM serves as the linear separator between two data points to identify two different classes in the multidimensional environment. SVM algorithms are in the binary format; so in the case of multi-class problem one must reduce the problem to a set of multiple binary classification problem.

Advantages

- Proposed support vector machine classifier is very useful when no clear idea (unbalanced data) about the data which is to be classified.

- Works well with even unstructured and semi structured data like text, Images and trees..
- It scales relatively well to high dimensional data.
- Computational complexity of classification is reduced

IV. ARCHITECTURE DIAGRAM

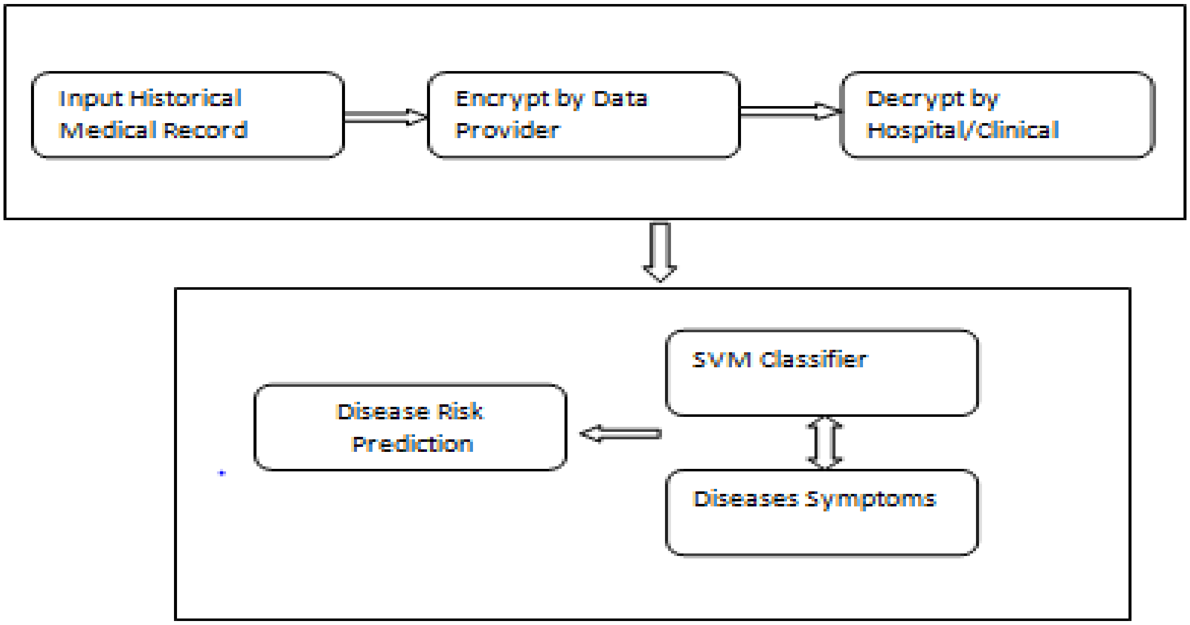

Fig.1 Overall System Architecture

V. MODULE DESCRIPTION

The proposed empirical system is designed and implemented with the following modules such as,

- Dataset Module
- Encryption Module
- Cloud Module
- Decryption Module
- SVM Classification Module

1) Dataset Module

The dataset is the training set created by a medical expert as a dataset to test the expert system, which was used to perform the presumptive diagnosis of diseases of the clinical detection system.

In this module, the data owner supplies the patient profile information and treatment history of the patients. The detail of the historical data of treatment includes the disease name, symptoms etc. These data is provided by the hospital admin as the data owner. The collected information is moved to the cloud server.

Data Provider can provide historical medical data that contain patients' symptoms and confirmed diseases, which are used for training support vector classifier. All these data are outsourced to cloud provider for storing.

2) Encryption Module

The hospital user submits the patient profile data and disease details of the patients to the cloud service provider. To maintain the privacy of the personal and medical data, the encryption module is developed. In this module, the given medical data about the symptoms and diseases were encrypted using the T-DES algorithm and moved to the cloud storage provider.

3) Cloud Module

The proposed privacy model, it tend to think that Data Provider is trustable which provides correct historical medical knowledge. The private party or the trusted third party processing unit is to be is considered as curious but honest, that is fascinated by data provider's individual medical knowledge and undiagnosed patient's medical knowledge but strictly follows the protocol within the system. The Trusted Authority is the indispensable entity, which is trusted by all entities involved in the system, which is in charge of distributing and managing all the private keys involve in the system. Undiagnosed patient is curious about the Classifier, whereas cloud provider is interested by all the other parties' knowledge within the system.

4) Decryption module

In the decryption module, the data owner data is decrypted to analyze the diseases risk of the patients by using the diseases symptoms. The trusted third party verifies the integrity constraints of the clinical laboratory user to outsource the patient private data including the disease details. Post successful verification of the user integrity constraints and data access rights by the trusted third party, the secret key is verified and the cloud storage provider invokes this decryption module to decode the encrypted data.

5) SVM Classification Module

A Support Vector Machine (SVM) performs classification by constructing an N-dimensional hyperplane that optimally separates the data into two categories. A predictor variable is called an attribute actor e, and a transformed attribute that is used to define the hyperplane is called a feature. The task of choosing the most appropriate representation is known as feature selection.

A set of features/attributes that describes one particular case is called a vector. So the goal of SVM modeling is to find the right hyperplane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and the other category are on the other side of the plane. The vectors near the hyperplane are the support vectors.

VI. CONCLUSION

In this paper proposed 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 support vector 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. Detailed privacy analysis ensures that patients' information is private and will not be leaked out during the disease diagnosis phase.

VII. FUTURE ENHANCEMENT

For Future enhancement, Identity Based Encryption is the best way to provide security for Public Health Record System. A new encryption technique with data auditing is used to verify the trustworthiness of third party auditor.

REFERENCES

[1] Y. Elmehdwi, B. K. Samanthula, andW. Jiang, "Secure k-nearest neighbor query over encrypted data in outsourced environments," in Proc. IEEE 30th Int. Conf. Data Eng., pp. 664–675, 2014.
[2] B. K. Samanthula, Y. Elmehdwi, G. Howser, and S. Madria, "A secure data sharing and query processing framework via federation of cloud computing," Inf. Syst., vol. 48, pp. 196–212, 2015.
[3] M. Kantarcıoglu, J. Vaidya, and C. Clifton, "Privacy preserving naïve bayes classifier for horizontally partitioned data," in Proc. IEEE Int. Conf. Workshop Privacy Preserving Data Min., 2003, pp. 3–9.
[4] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu, "Tools for privacy preserving distributed data mining," ACM SIGKDD Explorations Newslett., vol. 4, no. 2, pp. 28–34, 2002.
[5] X. Yi and Y. Zhang, "Privacy-preserving naive Bayes classification on distributed data via semi-trusted mixers," Inf. Syst., vol. 34, no. 3, pp. 371–380, 2009.
[6] M. Li, S. Yu, Y. Zheng, K. Ren, andW. Lou, "Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption," IEEE Trans. Parallel Distrib. Syst., vol. 24, no. 1, pp. 131–143, Jan. 2013.
[7] K. Lin and M. Chen, "On the design and analysis of the privacy-preserving SVM classifier," IEEE Trans. Knowl. Data Eng., vol. 23, no. 11, pp. 1704– 1717, Nov. 2011.
[8] H. Li, L. Xiong, L. Ohno-Machado, and X. Jiang, "Privacy preserving RBF kernel support vector machine," BioMed. Res. Int., vol. 2014, 2014.