Making Medication Choice from Heterogenous Data with Clinical Decision Support Framework

K. Tharani[1], R. Madhumeena[2]

[1,2]Deparment Of Computer Science and Engineering, Vellalar College of Engineering and Technology, Erode.
Email: tharani8888@gmail.com, rmadhumeena@gmail.com

Abstract — To keep pace with the developments in medical informatics, health medical data is being collected continually. But, owing to the diversity of its categories and sources, medical data has become highly complicated in many hospitals that it now needs Clinical Decision Support (CDS) system for its management. To effectively utilize the accumulating health data, we propose a CDS framework that can integrate heterogeneous health data from different sources, such as laboratory test results Diabetic basic information of patients, and health records into a consolidated representation of features of all patients. Using the electronic health Diabetic medical data so created, multi-label classification was employed to recommend a list of diseases and thus assist physicians in diagnosing or treating their patients' health issues more efficiently. Once the physician diagnoses the disease of a patient, the next step is to consider the likely complications of that disease, which can lead to more diseases. Previous studies reveal that correlations do exist among some diseases. Considering these correlations, a k-nearest neighbors' algorithm is improved for multi-label learning by using correlations among labels (CML-kNN). The CML-kNN algorithm first exploits the dependency between every two labels to update the origin label matrix and then performs multi-label learning to estimate the probabilities of labels by using the integrated and security features. Finally, it recommends the top N diseases to the physicians. To continue integrating textual and monitoring patient data to generate more comprehensive integrated features for each patient. The increasing diversity in data types calls for an appropriate method to decrease the number of integrated features for ensuring the efficiency of the clinical decision support system in Diabetic patient. Because of the scale of labels, the processing of improved multi-label algorithm will be a little slow. Therefore, a more appropriate and efficient method to correlate labels will have to be developed.

I. INTRODUCTION

Data mining, discovering of hidden predictive information from large data sets and it is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both. Data mining software is one of several analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

Data mining is also known as Knowledge Discovery in Database, refers to finding or "mining" knowledge from large amounts of data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. So, many people use the term "knowledge discovery in data" or KDD for data mining.

In Data mining, Knowledge extraction or discovery is done in seven sequential steps.

i. Data cleaning: This is the first step to eliminate noise data and irrelevant data from collected raw data.
ii. Data integration: At this step, various data sources are combined into meaningful and useful data.
iii. Data Selection: Here, data relevant to the analysis are retrieved from various resources.
iv. Data transformation: In this step, data is converted or consolidated into required forms for mining by performing different operations such as smoothing, normalization or aggregation.
v. Data Mining: At this step, various clever techniques and tools are applied in order to extract data pattern or rules.
vi. Pattern evaluation: At this step, Attractive patterns representing knowledge are identified based on given measures.

Knowledge representation: This is the last stage in which, visualization and knowledge representation techniques are used to help users to understand and interpret the data mining knowledge or result.

Clinical Decision Support System (CDSS) is an interactive Decision Support System (DSS) computer software, which is designed to assist physicians and other health professionals with decision making tasks, as determining diagnosis of patient data.

In the medical domain, a vast amount of knowledge is required even to solve seemingly simple problems. A physician is required to remember and apply knowledge of a vast array of documented disease presentations, diagnostic parameters, combination of drug therapies and guidelines. However the physician's cognitive abilities are restricted due to factors like multitasking, limited reasoning and memory capacity. Consequently, it is impossible for an unaided physician to make the right decision every time. Ironically, the increasing rate of information generated by medical advances has aggravated the physician's task.

The ideal decision making process contains the process of knowledge discovery. Many researchers consider data mining programs as a way to make decision making tools intelligent. The potential of computer based tools to address the medical decision making problems are realized half a century ago and several algorithms have been developed to construct Clinical Decision Support Systems for a variety of medical specialities.

II.OBJECTIVES

In m-healthcare social networks, the personal health information is always shared among the patients located in respective social communities suffering from the same disease for mutual support, and across distributed healthcare providers equipped with their own cloud servers for medical consultant. However, it also brings about a series of challenges, especially how to ensure the security and privacy of the patients' personal health information from various attacks in the wireless communication channel such as eavesdropping and tampering.

III.RELATED WORKS

In a m-healthcare system data confidentiality is much important but in existing system framework it is not enough for to only guarantee the data confidentiality of the patient's personal health information in the honest-but-curious cloud server model since the frequent communication between a patient and a professional physician can lead the adversary to conclude that the patient is suffering from a specific disease with a high probability. Unfortunately, the problem of how to protect both the patients' data confidentiality and identity privacy in the distributed m-healthcare Cloud Server scenario under the malicious model was left untouched. Once the physician diagnoses the disease of a patient, the next step is to consider the likely complications of that disease, which can lead to more diseases. Previous studies reveal that correlations do exist among some diseases. Considering these correlations, a k-nearest neighbors algorithm is improved for multi-label learning by using correlations among labels (CML-*k*NN). The CML-*k*NN algorithm first exploits the dependency between every two labels to update the origin label matrix and then performs multi-label learning to estimate the probabilities of labels by using the integrated features.

Drawbacks:

- In a tech-free system, the patients need to wait for a long time before they get to consult the doctor; besides, their repeated inquiries reduce diagnosis efficiency of their doctors. Also, inexperienced doctors may find it difficult to diagnose complicated illnesses
- The correlations among diseases are not considered by these systems.
- The correlations among diseases are not considered by these systems.
- Data confidentiality is low.
- Data redundancy is high.
- There is a violation in data security.

IV.METHODS

The main focus of this work is to build a clinical decision support framework for heterogeneous data sources (HDS CDS) for assisting doctors in diagnosing the diseases of their patients and treating them more efficiently. A novel framework is proposed for retrieving the most relevant information of patients from multiple data sources, such as laboratory test data, basic information of patients, symptoms of patients and electrocardiogram data, and for combining them to generate integrated features. A novel authorized accessible privacy model (AAPM) is established.

Patients can authorize physicians by setting an access tree supporting flexible threshold predicates. Then, based on it, by devising a new technique of attribute based designated verifier signature, a patient self-controllable multi-

level privacy preserving cooperative authentication scheme realizing three levels of security and privacy requirement in distributed m-healthcare cloud computing system is proposed. The directly authorized physicians, the indirectly authorized physicians and the unauthorized persons in medical consultation can respectively decipher the personal health information and/or verify patients' identities by satisfying the access tree with their own attribute sets. Considering the likely complications due to multiple medical diseases (conditions), k-nearest neighbors algorithm is proposed for multi-label learning, by using correlations among labels (CML-*k*NN) and for anticipating more potential diseases of a patient, so that a list of diseases can be recommended to the physician simultaneously. Using the laboratory test data and basic information of patients, a set of experiments of different multi-label learning methods were performed to confirm the effectiveness and practicality of the proposed framework.

- HDS CDS systems further improve the efficiency and effectiveness of CDS systems, especially by enhancing the wholeness of the system.
- HDS CDS systems collect and analyze healthcare data from diverse sources, rather than a single source.
- They suggest several correlative diseases by formulating a multi-label estimation model. Thus, HDS CDS system improves its performance in terms of comprehensiveness and accurate diagnosis. In the ongoing HDS CDS system, the patients initially offer information and, in return, they get information from both the system and the physician.
- As a result, the interaction among patients, physicians and system will be enhanced and information dissemination will improve
- M-healthcare system is fully controlled and secured with encryption standards.
- There is no data loss and data redundancy.
- System provides full protection for patient's data and their attributes.

V.CONCLUSION

In this paper, a novel authorized accessible privacy model (AAPM) and a patient self-controllable multi-level privacy preserving cooperative authentication scheme realizing three different levels of security and privacy requirement in the distributed m-healthcare cloud computing system are proposed, followed by the formal security proof and efficiency evaluations which illustrate our privacy preserving cooperative authentication scheme can resist various kinds of malicious attacks and far outperforms previous schemes in terms of storage, computational and communication overhead.

REFERENCES

[1] Femke Ongenae, Tom Dhaene, Filip De Turck, Dominique Benoit, Johan Decruyenaere, "Design of a Probabilistic Ontology-based Clinical Decision Support System for Classifying Temporal Patterns in the ICU: a Sepsis Case Study" IEEE, 2010.

[2] Jun Zhou, Xiaodong Lin, Xiaolei Dong, Zhenfu Cao, "PSMPA: Patient Self-controllable and Multi-level Privacy-preserving Cooperative Authentication in Distributed m-Healthcare Cloud Computing System" IEEE Transactions on Parallel and Distributed Systems, 2013.

[3] Linqi Song, William Hsu, Jie Xu, and Mihaela van der Schaar, "Using Contextual Learning to Improve Diagnostic Accuracy: Application in Breast Cancer Screening" IEEE Journal of Biomedical and Health Informatics, 2015.

[4] Min-Ling Zhang and Zhi-Hua Zhou," Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization" IEEE Transactions on Knowledge and Data Engineering, Vol.18, No.10, October 2006.

[5] OT Abdala, GD Clifford, M Saeed, A Reisner, G Moody, I Henry, RG Mark Harvard-MIT, "The Annotation Station: An Open-Source Technology for Annotating Large Biomedical Databases" Computers in Cardiology, 2004.

[6] Weider D. Yu, Choudhury Pratiksha, Sawant Swati, Sreenath Akhil, Medarametla Sarath, "A Modeling Approach to Big Data Based Recommendation Engine in Modern Health Care Environment" IEEE 39th Annual International Computers, Software & Applications Conference, 2015.

[7] William Van Woensel, Newres Al Haider, Patrice C. Roy, Ahmad Marwan Ahmad and Syed SR Abidi, "A Comparison of Mobile Rule Engines for Reasoning on Semantic Web Based Health Data" International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014.

[8] Xishui Pan, Xuezhong Zhou, Hongmei Song , Runshun Zhang, Tingting Zhang, " Enhanced Data Extraction, Transforming and Loading Processing for Traditional Chinese Medicine Clinical Data Warehouse" IEEE 14th International Conference on e-Health Networking, Applications and Service (Healthcom), 2012.

[9] Yu Wang, Peng-fei Li, Yu Tian, Jing-jing Ren, Jing-song Li, "A Shared Decision Making System for Diabetes Medication Choice Utilizing Electronic Health Record Data" IEEE Journal of Biomedical and Health Informatics, JBHI-00268-2016.

[10] Zhiwen Yu, Senior Member, IEEE, Peinan Luo, Jane You, Hau-San Wong, Hareton Leung, Si Wu, Jun Zhang, Senior Member, IEEE, Guoqiang Han, "Incremental Semi-supervised Clustering Ensemble for High Dimensional Data Clustering" IEEE transactions onknowledgeanddataengineering,2015.