OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

DEEP LEARNING APPROACH FOR CLINICAL RISK PREDICTION USING ELECTRONIC HEALTH RECORDS

Atul Balasaheb Bharte[1], Asst. Prof. V. S. Karwande[2]

Student, Department of CSE, EESGOI, Aurangabad, Maharashtra[1],

Assistant Professor, Department of CSE, EESGOI, Aurangabad, Maharashtra[2],

Atulbharate@gmail.com [1], Vijay.Karwande@Yahoo.Co.in[2]

Abstract: ***The broad acknowledgment of cloud based services in the healthcare sector has brought about practical and helpful trade of Personal Health Records (PHRs) among a few taking part elements of the e-Health systems. Nevertheless, putting away the secret health data to cloud servers is susceptible to revelation or theft and requires the improvement of approaches that guarantee the protection of the PHRs. Along these lines; we propose an approach called SeSPHR for secure sharing of the PHRs in the cloud. The SeSPHR scheme ensures patient-centric control on the PHRs and preserves the classification of the PHRs. The patients store the encrypted PHRs on the un-trusted in cloud servers and specifically grant access to various kinds of clients on various parts of the PHRs. A semi-trusted in proxy called Setup and Re-encryption Server (SRS) is acquainted with set up people in general/private key combines and to deliver the re-encryption keys. Besides, the strategy is secure against insider threats and furthermore authorizes a forward and in reverse access control. Moreover, we formally examine and check the working of SeSPHR strategy through the High Level Petri Nets (HLPN). Execution assessment in regards to time utilization demonstrates that the SeSPHR strategy can possibly be utilized for safely sharing the PHRs in the cloud.***

Keywords: *Cloud Computing, Personal Health Records, Privacy application.*

----------∴∴∴----------

I INTRODUCTION

Cloud computing has emerged as an important computing paradigm to offer pervasive and on demand availability of various resources in the form of hardware, software, infrastructure, and storage. The cloud computing also integrates various important entities of healthcare domains, such as patients, hospital staff including the doctors, nursing staff, pharmacies, and clinical laboratory personnel, insurance providers, and the service providers. Numerous methods have been employed to ensure the privacy of the PHRs stored on the cloud servers. The privacy preserving approaches make sure confidentiality, integrity, authenticity, accountability, and audit trial. Confidentiality ensures that the health information is entirely concealed to the unsanctioned parties, whereas integrity deals with maintaining the originality of the data, whether in transit or in cloud storage. Authenticity guarantees that the health-data is accessed by authorized entities only, whereas accountability refers to the fact that the data access policies must comply with the agreed upon procedures. We present a methodology called Secure Sharing of PHRs in the Cloud (SeSPHR) to administer the PHR access control mechanism managed by patients themselves. The methodology preserves the confidentiality of the PHRs by restricting the unauthorized users. Generally, there are two types of PHR users in the proposed approach, namely:

1. The patients or PHR owners and
2. The users of the PHRs other than the owners, such as the family members or friends of patients, doctors and physicians, health insurance companies representatives, pharmacists, and researchers.

The patients as the owners of the PHRs are permitted to upload the encrypted PHRs on the cloud by selectively granting the access to users over different portions of the PHRs. Each member of the group of users of later type is granted access to the PHRs by the PHR owners to a certain level depending upon the role of the user.

The levels of access granted to various categories of users are defined in the Access Control List (ACL) by the PHR owner. For example, the family members or friends of the patients may be given full access over the PHRs by the owner. Similarly, the representatives of the insurance company may only be able to access the portions of PHRs containing information about the health insurance claims while the other confidential medical information, such as medical history of the patient is restricted for such users.

Recently, the increased availability of electronic health records (EHR) has demonstrated great potential to improve the performance of clinical risk prediction. EHRs regularly record various care and treatment behaviors, e.g., procedures, diagnoses, and laboratory tests and measurements, etc., of patients during their hospitalization within the context of large healthcare systems this captures the characteristics of Heterogeneous populations of patients receiving care in their current clinical setting.

There is an excellent opportunity to develop more accurate risk prediction models from EHRs. Effective prediction of clinical risks of ACS patients via their heterogeneous EHR data is still an intricate problem and remains a major challenge for healthcare management, mainly due to high clinical complexity and the natural heterogeneity in EHR data. Therefore, one of the most important tasks in clinical risk prediction is to develop robust prediction models that can effectively handle high dimensional heterogeneous EHR data and accurately classify different clinical risks levels based on the acquired EHR data.

II LITERATURE SURVEY

2.1 Minas A. Karaolis, Joseph A. Moutiris, Demetra Hadjipanayi, Constantinos S. Pattichis, "Assessment of the Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees", IEEE Transactions on Information Technology in Biomedicine, 2010.

Coronary heart disease (CHD) is one of the major causes of disability in adults as well as one of the main causes of death in the developed countries. Although significant progress has been made in the diagnosis and treatment of CHD, further investigation is still needed. The objective of this study was to develop a data-mining system for the assessment of heart event-related risk factors targeting in the reduction of CHD events. The risk factors investigated were:
1) before the event: a) nonmodifiable-age, sex, and family history for premature CHD, b) modifiable-smoking before the event, history of hypertension, and history of diabetes; and 2) after the event: modifiable-smoking after the event, systolic blood pressure, diastolic blood pressure, total cholesterol, high-density lipoprotein, low-density lipoprotein, triglycerides, and glucose. The events investigated were: myocardial infarction (MI), percutaneous coronary intervention (PCI), and coronary artery bypass graft surgery (CABG). A total of 528 cases were collected from the Paphos district in Cyprus, most of them with more than one event. Data-mining analysis was carried out using the C4.5 decision tree algorithm for the aforementioned three events using five different splitting criteria. The most important risk factors, as extracted from the classification rules analysis were: 1) for MI, age, smoking, and history of hypertension; 2) for PCI, family history, history of hypertension, and history of diabetes; and 3) for CABG, age, history of hypertension, and smoking. Most of these risk factors were also extracted by other investigators. The highest percentages of correct classifications achieved were 66%, 75%, and 75% for the MI, PCI, and CABG models, respectively. It is anticipated that data mining could help in the identification of high and low risk subgroups of subjects, a decisive factor for the selection of therapy, i.e., medical or surgical.

2.2. Muxuan Liang, Zhizhong Li, Ting Chen, Jianyang Zeng, "Integrative Data Analysis of Multi-Platform Cancer Data with a Multimodal Deep Learning Approach", IEEE/ACM Transactions on Computational Biology and Bioinformatic, 2015.

Identification of cancer subtypes plays an important role in revealing useful insights into disease pathogenesis and advancing personalized therapy. The recent development of high-throughput sequencing technologies has enabled the rapid collection of multi-platform genomic data (e.g., gene expression, miRNA expression, and DNA methylation) for the same set of tumor samples. Although numerous integrative clustering approaches have been developed to analyze cancer data, few of them are particularly designed to exploit both deep intrinsic statistical properties of each input modality and complex cross-modality correlations among multi-platform input data. In this paper, we propose a new machine learning model, called multimodal deep belief network (DBN), to cluster cancer patients from multi-platform observation data.

In our integrative clustering framework, relationships among inherent features of each single modality are first encoded into multiple layers of hidden variables, and then a joint latent model is employed to fuse common features derived from multiple input modalities. A practical learning algorithm, called contrastive divergence (CD), is applied to infer the parameters of our multimodal DBN model in an unsupervised manner. Tests on two available cancer datasets show that our integrative data analysis approach can effectively extract a unified representation of latent features to capture both intra- and cross-modality correlations, and identify meaningful disease subtypes from multi-platform cancer data. In addition, our approach can identify key genes and miRNAs that may play distinct roles in the pathogenesis of different cancer subtypes. Among those key miRNAs, we found that the expression level of miR-29a is highly correlated with survival time in ovarian cancer patients.

2.3. Assad Abbas, Samee U. Khan, Senior Member, "A Review on the Stateof- the-Art Privacy Preserving Approaches in the e-Health Clouds", IEEE Journal of Biomedical and Health Informatics, 2014.

Cloud computing is emerging as a new computing paradigm in the healthcare sector besides other business domains. Large numbers of health organizations have started shifting the electronic health information to the cloud environment. Introducing the cloud services in the health sector not only facilitates the exchange of electronic medical records among the hospitals and clinics, but also enables the cloud to act as a medical record storage center. Moreover, shifting to the cloud environment relieves the healthcare organizations of the tedious tasks of infrastructure management and also minimizes development and maintenance costs. Nonetheless, storing the patient health data in the third-party servers also entails serious threats to data privacy. Because of probable disclosure of medical records stored and exchanged in the cloud, the patients' privacy concerns should essentially be considered when designing the security and privacy mechanisms. Various approaches have

been used to preserve the privacy of the health information in the cloud environment. This survey aims to encompass the state-of-the-art privacy-preserving approaches employed in the e-Health clouds. Moreover, the privacy-preserving approaches are classified into cryptographic and noncryptographic approaches and taxonomy of the approaches is also presented. Furthermore, the strengths and weaknesses of the presented approaches are reported and some open issues are highlighted.

2.4. M. H. Au, T. H. Yuen, J. K. Liu, W. Susilo, X. Huang, Y. Xiang, and Z. L. Jiang, "A General Framework for Secure Sharing of Personal Health Records in Cloud System", Journal of Computer and System Sciences, 2017.

In this paper, Author provided an affirmative answer to this problem by presenting a general framework for secure sharing of PHRs. This system enables patients to securely store and share their PHR in the cloud server (for example, to their careers), and furthermore the treating doctors can refer the patients' medical record to specialists for research purposes, whenever they are required, while ensuring that the patients' information remain private. This system also supports cross domain operations (e.g., with different countries regulations).

2.5. David Daglish and Norm Archer, "Electronic Personal Health Record Systems: A Brief Review of Privacy, Security, and Architectural Issues", IEEE 9World Congress on Privacy, Security, Trust and the Management of e-Business, 2009.

Electronic personal health records (PHRs) are beginning to receive widespread attention as a tool for consumers. Such systems may be used by individuals to input data and to access information from a variety of sources (e.g. family physicians), thus improving their understanding of the state of their health and helping to manage their own healthcare better. The main source of information for PHRs is normally the patient's physician, supplemented by patient input and other sources of information such as prescriptions and lab test results, as well as institutional inputs from hospitals and other facilities. The architecture of such a system must be such that patients can access all the useful information that is relevant to their medical history in a form that is understandable to them, while at the same time protecting against unauthorized access. This paper addresses design and architectural issues of PHR systems, and focuses on privacy and security issues which must be addressed carefully if PHRs are to become generally acceptable to consumers.

2.6. S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable and Fine-grained Data Access Control in Cloud Computing", in Proceedings of the IEEE INFOCOM, March 2010.

Cloud computing is an emerging computing paradigm in which resources of the computing infrastructure are provided as services over the Internet. As promising as it is, this paradigm also brings forth many new challenges for data security and access control when users outsource sensitive data for sharing on cloud servers, which are not within the same trusted domain as data owners. To keep sensitive user data confidential against untrusted servers, existing solutions usually apply cryptographic methods by disclosing data decryption keys only to authorized users. However, in doing so, these solutions inevitably introduce a heavy computation overhead on the data owner for key distribution and data management when fine-grained data access control is desired, and thus do not scale well.

The problem of simultaneously achieving fine-grainedness, scalability, and data confidentiality of access control actually still remains unresolved. This paper addresses this challenging open issue by, on one hand, defining and enforcing access policies based on data attributes, and, on the other hand, allowing the data owner to delegate most of the computation tasks involved in fine-grained data access control to untrusted cloud servers without disclosing the underlying data contents. We achieve this goal by exploiting and uniquely combining techniques of attribute-based encryption (ABE), proxy re-encryption, and lazy re-encryption. Our proposed scheme also has salient properties of user access privilege confidentiality and user secret key accountability. Extensive analysis shows that our proposed scheme is highly efficient and provably secure under existing security models.

III SYSTEM ARCHITECTURE

Propose clinical risk prediction model using deep learning for analyzing a large volume of multidimensional heterogeneous EHR data. We firstly formulate the proposed problem, and then propose our learning schema for clinical risk prediction. Moreover, we present the approach to extract potential informative risk factors via the reconstruction learning strategy. Proposes a novel approach for clinical risk prediction of ACS based on deep learning. Our proposed model can learn more discriminative patient feature representations and thus improve the performance of clinical risk prediction.

To evaluate the proposed approach, extensive experiments using an EHR dataset consisting of 3,464 ACS patient samples and collected from the Cardiology department of the Chinese PLA General Hospital were carried out. Specifically, we extracted patient's information from admission records to construct clinical risk prediction models for ACS patients. As shown in Fig.1, an admission record sample contains valuable patient information such as demographics, medical history, physical examination results, lab test and specific inspection, and the first primary diagnosis and common comorbidities, etc. In clinical practice, physicians often refer to the admission record of the ACS patient to determine his or her clinical risk.

We propose to utilize admission records to construct our clinical risk prediction model and learn informative ACS risk factors for helping physicians predict clinical risks for ACS patients at the early stage of their hospitalizations, so as to make the practice better for the care of individual patients. The proposed regularized SDAE keeps in memory of the characteristics of patient risk information during learning, and thus it has the ability to enforce the reconstructed feature representations within the same risk level to be as close as possible and the reconstructed feature representations between different risk levels to be kept distant as much as possible. We

apply this regularized SDAE to pre-train a clinical risk prediction model from EHR data.

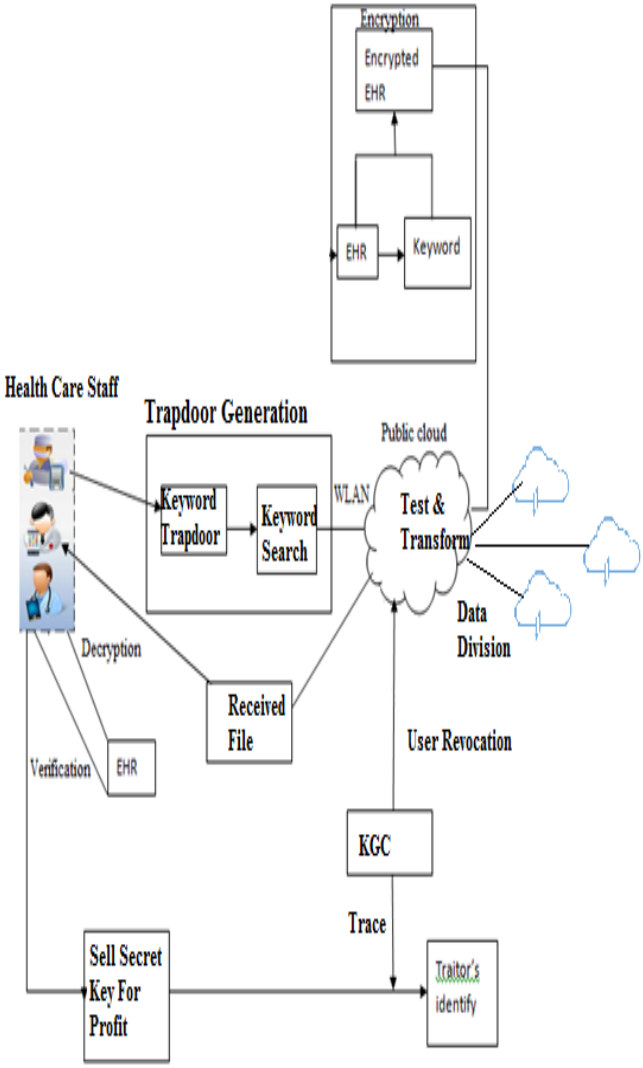

Figure 1: Architecture Diagram

IV RESULT

In proposed system, data is stored on cloud. Before uploading files on cloud, files encrypted and then stores on cloud. While storing files on cloud, it will take some time to write files on cloud. In experiment, file size considered in kb, as file size increase required time to uploading increases exponentially

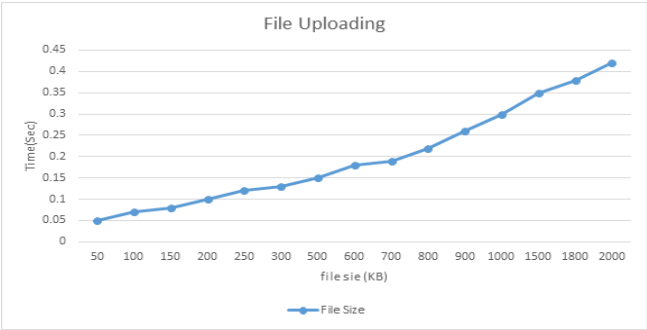

Figure 2: File Uploading

Module 1: Data Owner (Patient)

Data Owner Module contains patient's health information, which is controlled by the patients themselves. The PHRs permit patients to manage the information, such as demographics, diagnosis, treatments, monitoring, and self-care. The PHRs are different from the Electronic Health Records (EHRs) in the sense that the EHRs
are managed by the health organizations and contain the information entered by the doctors and the hospital staff instead of patients.

Module 2: Data User

Doctors, nursing staff, pharmacies, clinical laboratory personnel, insurance providers, and the service providers is the data users in Health network. Each data user has a set of attributes, such as affiliation, department and type of health care staff, and is authorized to search on encrypted EHRs based on his set of attributes. In SeSPHR, a data uses resource-limited terminals to generate secret keys and conduct the information retrieval operation. The secret keys are sent to the public cloud via wireless channel and the retrieved EHR files are returned. Then, the data user decrypts the EHR files and verifies the correctness of decryption.

Module 3: Public Cloud

The public cloud has almost unlimited storage and computing power to undertake the PHR remote storage task and respond on data retrieval requests. Lightweight test algorithm is designed in our proposed system to improve performance.

Module 4: Key Generation Center(KGC)

KGC generates public parameters for the entire system and distributes secret keys to data users. A data users set of attributes is embedded in his secret key to realize access control. If a traitor sells his secret key for financial gain, the KGC is able to trace the identity of the malicious user and revoke his secret key.

V CONCLUSION

We proposed a procedure to safely store and transmission of the PHRs to the authorized elements in the cloud. The strategy preserves the security of the PHRs and authorizes a patient-driven access control to various segments of the PHRs on the access provided by the patients. We executed a fine-grained access control technique so that even the valid system clients can't get to those segments of the PHR for which they are not authorized. The PHR owners

store the encrypted information on the cloud and just the approved users having valid re-encryption keys issued by a semi trusted authority can decrypt the PHRs. The job of the semi-trusted authority is to produce and store the public/private key sets for the clients in the system. The performance Evaluation was done on the based on time required to generate keys, encryption and decryption tasks, and turnaround time. The trial results display the reasonability of the SeSPHR system to secure share the PHRs in the cloud environment.

REFERENCES

[1] Minas A. Karaolis, Joseph A. Moutiris, Demetra Hadjipanayi, Constantinos S.Pattichis, "Assessment of the

Risk Factors of Coronary Heart Events Based on Data Mining With Decision Trees", IEEE Transactions on Information Technology in Biomedicine, 2010.

[2] Muxuan Liang, Zhizhong Li, Ting Chen, Jianyang Zeng, "Integrative Data Analysis of Multi-Platform Cancer Data with a Multimodal Deep Learning Approach", IEEE/ACM Transactions on Computational Biology and Bioinformatic, 2015.

[3] Assad Abbas, Samee U. Khan, Senior Member, "A Review on the State-of-the-Art Privacy Preserving Approaches in the e-Health Clouds", IEEE Journal of Biomedical and Health Informatics, 2014.

[4] M. H. Au, T. H. Yuen, J. K. Liu, W. Susilo, X. Huang, Y. Xiang, and Z. L. Jiang, "A General Framework for Secure Sharing of Personal Health Records in Cloud System", Journal of Computer and System Sciences, 2017.

[5] David Daglish and Norm Archer, "Electronic Personal Health Record Systems: A Brief Review of Privacy, Security, and Architectural Issues", IEEE 9 World Congress on Privacy, Security, Trust and the Management of e-Business, 2009.

[6] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable and Finegrained Data Access Control in Cloud Computing", in Proceedings of the IEEE INFOCOM, March 2010.

[7] K. Gai, M. Qiu, Z. Xiong, and M. Liu, "Privacy-preserving Multi-channel Communication in Edge-of-Things", Future Generation Computer Systems, 2018.

[8] K. Gai, M. Qiu, and X. Sun, "A Survey on FinTech", Journal of Network and Computer Applications, 2017.

[9] A. Abbas, K. Bilal, L. Zhang, and S. U. Khan, "A Cloud based Health Insurance Plan Recommendation System: A User Centered Approach, Future Generation Computer Systems, 2015.

[10] J. Li, "Electronic Personal Health Records and the Question of Privacy", Computers,
2013.