Pak. J. Biotechnol. Vol. 14 (Special Issue II) Pp. 323 -327 (2017) Leelavathy. S., S. Selva Brunda

A CLOUD BASED SECURED FRAMEWORK FOR SMART MEDICAL DIAGONSIS: A SURVEY

Leelavathy. S[1]., S. Selva Brunda[2]

Bharathiar University, Coimbatore, India. Cheran College of Engineering, Karur
leelasiva5307@gmail.com, brindhaselva@yahoo.com

ABSTRACT

Enormous amounts of medical data are being collected by several well-developed HIS (Hospital Information Systems) in the form of patient records in hospitals. The hidden patterns and relationships contained in this data are identified using various data mining techniques which has drawn increasing world-wide attention in the recent years. As a result of which, a good number of MDSS (Medical Decision Support Systems) have been developed. These are computer systems designed to assist physicians or other healthcare professionals in making clinical decisions for the given patient's symptoms and medical history. Inspired by the existing systems, the model proposed in this paper aims at making a knowledge sharing collaborative platform for doctors which serves as a realistic and effective Medical Decision Support System. It is important because it provides vital information from different data sources. It has several challenges like scalability, response time, heterogeneous data formats. This paper does a detailed study of different challenges in understanding about medical information that can be provided to users in a better manner. A framework named Intelligent Healthcare Framework (ISHF) is being proposed in this paper.

keywords—Medical decision support system (MDSS), diagnosis, knowledge discovery and symptoms

I. INTRODUCTION

Getting to the right diagnosis is the critical step of beginning the treatment for the patient. A world-famous doctor, Leana Wen, said "We know that at least 80% of diagnoses can be made based on the history alone. If the focus is on the diagnosis right, rather than blindly ordering tests, it would actually save both the patient and the doctor a lot of time. Not to mention that it would eliminate unnecessary side effects such as radiation of a CAT scan, and the problems that come with delay to diagnosis. More importantly, a patient's health is too important to get it wrong."

So, to assist doctors in making better diagnosis, a variety of Medical Decision Support Systems [1-3] have been designed using various data mining techniques like Support Vector Machine [4], Bayesian Classifiers [4] and even using a combination of techniques – Differential Diagnosis method [5], LAMSTAR and k-NN together [6] to improve accuracy and efficiency of the system. These data mining techniques learn the patterns and relationships between various attributes in the medical dataset which consists of various patients' symptoms, medical history and diagnosis given by the doctor. So, when a new patient's symptoms and medical history is given as an input, these systems give the percentage of occurrence of most probable diseases.

In reality, the field of medicine can't be compared with any field like stock marketing to take an important decision based on the resultant figures produced by a system after analysis of past trends. In case of stock market analysis, figures are worth a thousand words. But, when it comes to medical diagnosis, just the figures (probability of occurrence of most probable diseases) obtained from a decision support system wouldn't be sufficient for the challenging task of getting to the right diagnosis.

The very strength of the field of medicine is 'knowledge and experience' of doctors. So, in view of the above-mentioned factors, we would like to propose a knowledge sharing collaborative medical decision support system which speaks more than figures. In this system, we give special importance to knowledge and experience of doctors in their fields of specialisation.

From our survey, we also realized that many doctors would like to get suggestions from other doctors in their field of specialization while dealing with complicated cases of patients. They hardly get time to refer to the books and other sources. This web-based system would save a lot of their time by giving the diagnosis of similar complicated cases diagnosed by other doctors for some patients somewhere. Besides seeing the other similar patients' symptoms, medical history and doctors' diagnosis for them, he/she can also see the 'final effective diagnosis' which is determined by taking patient's feedback, doctor's diagnosis, doctor's experience in the field of specialization and a lot more.

II. RELATED WORK

A. **Dataset Detail:** A Medical Dataset comprises of symptoms and medical history of the patient and the corresponding diagnosis for the disease given by the doctor. As an example, for a sample medical dataset, below are attributes of a 'Heart Disease Data Set' obtained from UCI Machine Learning Repository.

The UCI ML Repository stated that "The Cleveland database is the only one that has been used by ML researchers to this date." This dataset consists of 14 attributes. They are:

1. Age: in years
2. Sex: (1=male; 0=female)
3. Chest pain type: -- Value 1: typical angina
   -- Value 2: atypical angina
   -- Value 3: non-anginal pain
   -- Value 4: asymptomatic

4. Resting blood pressure (in mm Hg on admission to the hospital)
5. Serum cholesterol in mg/dl
6. Fasting blood sugar, -- Value 1: (> 120 mg/dl) -- Value 0: otherwise
7. Restecg: resting electrocardiographic results -- Value 0: normal

-- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria

8. Thalach: maximum heart rate achieved
9. Exang: exercise induced angina (1 = yes; 0 = no)
10. Old peak: ST depression induced by exercise
11. Slope: the slope of the peak exercise ST segment -- Value 1: upsloping, -- Value 2: flat -- Value 3: down sloping
12. CA: number of major vessels (0-3) colored by fluoroscopy
13. Thal: -- Value 3 = normal,

-- Value 6 = fixed defect, -- Value 7 = reversible defect

14. Diagnosis of heart disease (angiographic disease status), - Value 0: < 50% diameter narrowing -- Values [1-4]: > 50% diameter narrowing

Note that the last field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4.

B. **Health Care:** This section presents study on different details of health care. The aspects considered are functional aspects of health care, technology, detail maintenance.

Table 1: Health care Trend and Diagnosis procedure

S.No	Title of the paper	Authors	Journal/Conference	Key points learnt
1	Introduction to the Special Issue on Advances in Clinical and Health-Care Knowledge Management	Rajeev K. Bali David Dagan Feng Frada Burstein Ashish N. Dwivedi	IEEE Transactions on Information Technology in Biomedicine 9(2): 2005	Importance of health-care management concepts such as Evidence-Based Medicine (EBM) and Knowledge Management (KM) which include aspects of clinical information management through data mining, artificial intelligence and many other techniques.
2	Knowledge Discovery in Medical Systems Using Differential Diagnosis, LAMSTAR, and k-NN	Rahul Isola Rebeck Carvalhod Amiya Kumar Tripathy.	IEEE Transactions on Information Technology in Biomedicine16(6): 2012	Focuses on computing the probability of occurrence of a particular ailment from the medical data by mining it using unique algorithm that increases accuracy of such diagnosis by combining the key points of neural networks, Large Memory Storage, Retrieval, k-NN, and differential diagnosis all integrated into algorithm
3	Discover the Expert: Context-Adaptive Expert Selection for Medical Diagnosis	C. Tekin, Onur atan, Mihaela Van Der Schaar	IEEE Transactions on Emerging topics in Computing 3(2): 2015	Aims to optimize clinical workflows by personalizing the match of (new patient) cases with the appropriate diagnostic expertise through a Clinical Decision Support Systems (CDSS) to best choose the domain expert who specializes in similar types of cases.
4	Data Mining in Hospital Information System for Hospital Management	Shusaku Tsumoto Shoji Hirano	IEEE conference	Shows application of exploratory data analysis techniques to data extracted from Hospital Information System (HIS), which can be viewed as a first step to data mining for HIS. The results show that the results are very useful for Hospital Management and that the reuse of stored data will give a powerful tool to support a long-period management of a hospital.
5	Specialised Tools for Automating Data Mining for Hospital Management.	J. Alapont, A. Bella-Sanjuan, C. Ferri, J. Hernandez, J. D. L lopis, M.J. Ramírez.	Supported by the Valencian Government through an IMPIVA project	Covers all the stages in the process of Knowledge Discovery from Databases (KDD): data cleansing, extraction and integration from the HIS (Hospital information system) and external data, construction of tasks and minable views, model generation, and finally a module to carry out and interpret their predictions.

C. Cloud Based services

Khoshkbarforoushhaa et al., (6) analyzed the elasticity aspects of cloud. A cloud-hosted SDAF typically consists of three layers: data ingestion, analytics,and storage. The data ingestion layer accepts data from multiple sources such as online services or back-end system logs. The data analytics layer consists of many platforms including stream/batch processing systems, and scalable machine learning frameworks that ease implementation of data analytics use-cases such as collaborative filtering and sentiment analysis. The ingestion and analytics layers make use of different databases during execution and where required persist the data in the storage layer. A architecture similar to Amazon is proposed. However, it can be extended to Health care application.

Fazia et al., (7) analyzed the open issues in cloud. The different micro services are identified. The proposed

microservice based architecture is suitable for health care as there are different foreign cloud services operating in the same domain.

Fazia et al., (7) analyzed the remote health care aspects. The different sensors are also identified and the sensor based challenges are identified. The paper is important as it speaks about different challenges in terms of users and technology.

Alqahtani et al., (8) analyzed the Qos aspects of Internet of Things. A architecture is proposed that has the characteristics required for cross functional aspects of communication in real time.

Khoshkbarforoushha et al., (9) analyzed the aspects of how queries are complex in nature and how they could be handled using Density networks. The nature of data in terms of streaming is also considered. Wang et al., (12) analyzed the different tools and technologies for cloud.

Song et al., (10) analyzed the different challenges in terms of android application. It understands the manifest files in terms of understanding the MD5 types.

Madhukar and Ragunathan (11) analyzed the cloud scheduling algorithm. The different jobs are processed in queues. The queue management is dynamic and is handled in terms of batches.

Lu et al., (13) analyzed how Azure serices can provide a solution on cloud. The functional aspects, patient level diagnostics are however presented not to a detailed level.

Inference

a. The query handled proposed could be automated for health care

b. The cloud based issues for health care could be studied in detail

c. Deploying cloud on Android platform requires understanding about MD5.

d. Different user requirements in health care can be processed as batches in queue.

D. Search Algorithms

Navimipour and Milani (17) analyzed the importance of search algorithms on cloud. A modified search algorithm based on Cuckoo search algorithm is proposed. The execution time is taken as a metric for measuring performance.

Angelo and Giaccari (18) analyzed the different comparisons between kdd and btree algorithms on cloud. The different comparisons between the algorithms in terms of time and search space is being evaluated.

Makkar and Dalal (19) suggested a ranked keyword search for searching data on the datastore. The details provided are based on understanding how the ranked keyword search will provide correct results in terms of search space and time.

Swapna et al., (20) analyzed the architecture aspects of large network communication. The paper discusses how algorithms like LSH can provide a better solution for searching.

Inference

1. The search algorithms suitably applied will reduce the search time.
2. LSH algorithm is a suitable algorithm for searching data and can be applied for health care.

E. Statistical Techniques

Statistical approach involves training and test phase. The input data is understood in terms of features. The features are represented and modeled as feature vectors. The statistical parameters could be tuned for better learning.

Ani1 et al., (14) made a detailed study on different statistical approaches and classifiers. The detailed ideas on how the classifier could be combined is also presented in the paper. The study provides directions for unsupervised learning in complete manner based on appropriate feature selection and prediction.

Olivier et al., (15) suggested a variation to statistics in terms of optimization approaches. The deviation is considered as a key factor for understanding how the experiments can be carried out using optimization theory.

Michael et al., (16) described the different approaches for understanding how massively large data can be analyzed. The paper suggests how Biglm algorithm can become a solution for massively large data.

Inference

1. The health care data is a type of massive dataset
2. The data is huge and Biglm is a suitable algorithm for predictive analytics

III. Outcomes of the survey

As a outcome of the survey a suitable framework is being proposed as below:

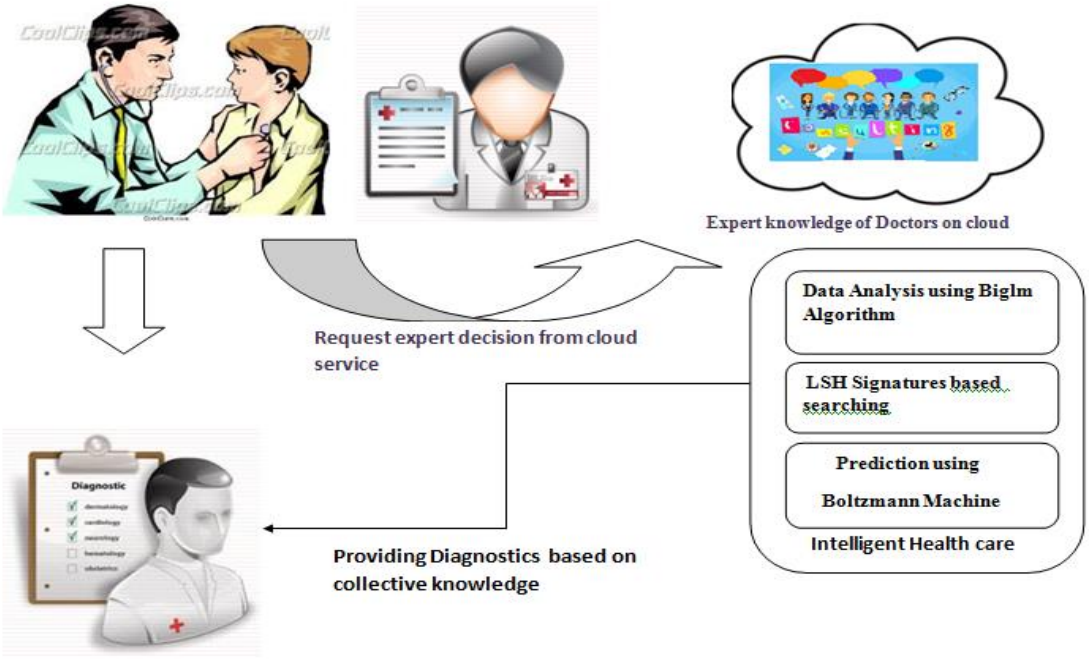

Figure 3. Architecture Diagram of the Intelligent Health care Framework(ISCF)

The proposed framework works as follows:

1. The patient details are collected along with other tests and are recorded to the expert database
2. The patient recorded details regarding health is being pushed to the cloud for expert advice.
3. The framework understands the nature of the data using Biglm algorithm for prediction.
4. The data retrieval will become efficient using LSH signature algorithm
5. The prediction is completed using Boltzmann Machine.
6. The retrieved results are sent to the doctor for providing diagnostic report.

III. Conclusion

The study completed in terms of understanding how the health care can be done at geographical level under one umbrella provides challenges in terms of (i) securety, privacy, and integrity of confidential medical data (ii) QoS in ensuring right information and right time (iii) easy integration allowing for homogenous communication (iv) A single layer of encapsulation for providing multiple services. The framework proposed will allow knowledge-driven optimized resource provisioning that adapts to uncertain data streams/volumes, the number of users and varying resource, and workload unpredictability. The framework also provides support for understanding the prediction in terms of different inputs from several specialists around the world. The survey completed will become challenging as it involves many inputs from specialist in the concerned field.

IV. References

1.Rajeev K.B., D.D. Feng, F. Burstein, A.N. Dwivedi, Introduction to the Special Issue on Advances in Clinical and Health-Care Knowledge Management. IEEE Transactions on Information Technology in Bio-medicine 9(2): (2005)

2. Rahul Isola, R. Carvalho, A.K. Tripathy, Knowledge Discovery in Medical Systems Using Differential Diagnosis, Lamstar, and k-nn. IEEE Transactions on Information Technology in Biomedicine 16(6): (2012)

3. Tekin, C. O. Atan, M. Van Der Schaar, Discover the Expert: Context-Adaptive Expert Selection for Medical Diagnosis. IEEE Transactions on Emerging Topics in Computing 3(2): 220-234 (2015)

4. Shusaku Tsumoto Shoji Hirano, Data Mining in Hospital Information System for Hospital Management, IEEE Conference (2014)

5. Alapont, J., A. Bella-Sanjuan, C. Ferri, J.Hernandez, J.D. Llopis, M.J. Ramírez, Specialised Tools for Automating Data Mining for Hospital Management, Supported by the Valencian Government through an IMPIVA project.

6. Khoshkbarforoushhaa, A., A. Khosravianc and R. Ranjan, Elasticity Management of Streaming Data Analytics Flows on Clouds. Journal of Computer System Sciences In Press (2016)

7. Fazia, M., A. Celesti, R. Ranjan, L. Chen, C. Liu, and M. Villari, Open Issues in Scheduling Microservices in the Cloud, IEEE Cloud Computing, IEEE Computer Society (2016)

8. Alqahtani, A., E. Solaiman, R. Buyya and R. Ranjan, End-to-End QoS Specification and Monitoring in the Internet of Things. Newsletter, IEEE Technical Committee on Cybernetics for Cyber-Physical Systems 1(2): (2016).
9. Khoshkbarforoushha, A., R. Ranjan, R. Gaire, E. Abbasnejad, L. Wang and Albert Y. Zomaya, Distribution Based Workload Modelling of Continuous Queries in Clouds. IEEE Transactions on Emerging Topics in Computing, IEEE Computer Society (2016).
10. Song, J., C. Han, K. Wang, J. Zhao, R. Ranjan and L. Wang, An Integrated Static Detection and Analysis Framework for Android, Journal of Pervasive and Mobile Computing 32: 15-25 (2016).
11. Madhukar, E. and T. Ragunathan, Efficient Scheduling Algorithm for Cloud, Elesiver Procedia-Big Data, Cloud and Computing Challenges 50: 353-356 (2015).
12. Wang, L., R. Ranjan, J. Kolodziej, A. Zomaya and L. Alem, Software Tools and Techniques for Big Data Computing in Healthcare Clouds. Future Generation Computer Systems Journal 43-44: 38-39 (2015).
13. Lu, S., R. Ranjan and P. Strazdins, Reporting an Experience on Design and Implementation of e-Health Systems on Azure Cloud. Journal of Concurrency and Computation: Practice and Experience (CCPE) (2014)
14. Ani1 K. Jain, Robert P.W. Duin, Jianchang Mao, Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1): (2000)
15. Olivier François, Christian Lavergne, Design of Evolutionary Algorithms—A Statistical Perspective. IEEE Transactions on Evolutionary Computation 5(2): 29-149 (2001).
16. Michael J. Kane, John W. Emerson, Stephen Weston, Scalable Strategies for Computing with Massive Data. Journal of Statistical Software 55 (14): 1 - 20 (2013)
17. Navimipour N.F. and F.S. Milani, Task Scheduling in the Cloud Computing Based on the Cuckoo Search Algorithm. International Journal of Modeling and Optimization 5(1): 44-47 (2015)
18. Angelo1, L.D. and L. Giaccari, An efficient algorithm for the nearest neighborhood search for point clouds. International Journal of Computer Science Issues 8(5): (2011) .
19. Makkar V. and S. Dalal, Ranked Keyword Search in Cloud Computing: An Innovative Approach. International Journal of Computational Engineering Research 3(6): (2013).
20. Swapna, P., D. Hari Krishna, Using State-of the-art Algorithm Fast and Efficient Search over Encrypted Data. IJDCST 1(6): 81-86 (2013)